Polytechnic Institute of New York

LEVEL II

12

POLY-EE/CS-81-002
June 1981

DETECTION OF MOVING TARGETS BY ADAPTIVE CLUTTER SUPPRESSION

by

A. Papoulis

FINAL REPORT

Contract No. N00014-76-C-0144
ARPA Order No. 2684
Scientific Officer; F.W.Quelle

October 1978 thru June 1981

Sponsored by
Defense Advanced Research Projects Agency

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO. AD-A102273

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Detection of Moving Targets by Adaptive Clutter Suppression.

5. TYPE OF REPORT & PERIOD COVERED
FINAL REPORT
Oct 1978 — June 1981

6. PERFORMING ORG. REPORT NUMBER
POLY-EE/CS-81-002

7. AUTHOR(s)
A. Papoulis

8. CONTRACT OR GRANT NUMBER(s)
N00014-76-C-0144
ARPA Order-2684

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Polytechnic Institute of New York
Route #110
Farmingdale, New York 11758

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
DARPA

12. REPORT DATE
June 81

13. NUMBER OF PAGES
62

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC
ELECTE
JUL 31 1981
S C

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
clutter suppression, image restoration, moving targets, spectral estimation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
An adaptive filtering method is developed for detecting a moving target in the presence of strong interference. The motion of the target is utilized to create a high frequency component in the observed signal. The filtering is performed in terms of the running transform of the data and the local statistics of the interference. Other problems considered include spectral estimation and image restoration by extrapolation.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified 411284
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Final Report

In this year's effort, research focussed on four areas.

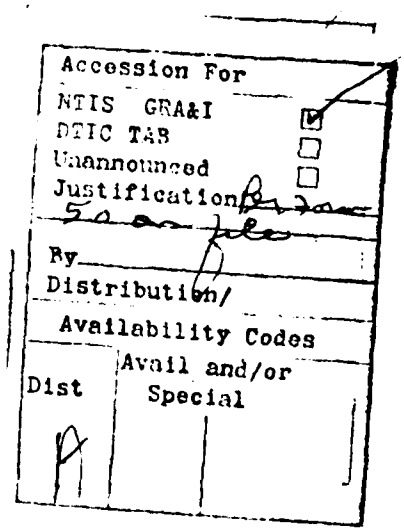
1. Adaptive Filtering and Estimation.

Work on the theory of running FFT's continued with emphasis on problems in adaptive filtering. Results were presented in the following paper:

"Adaptive Frequency Domain Estimators"
IEEE International Symposium on Information Theory, Grignano, Italy, 1979.

The method was applied to the problem of detecting a moving target in the presence of strong clutter. The filtering was based not on global but on local spectral properties of the clutter determined adaptively with threshold and other techniques. Results were presented in the following paper:

"Adaptive Clutter Suppression"
Seventh DARPA Strategic Space Symposium, Naval Post Graduate School, Monterey, California, 1980.



29. ADAPTIVE CLUTTER SUPPRESSION*

By: A. Papoulis K. Huang and Ch. Chamzas†

Abstract--A method of target detection is presented based on the determination of the local spectral properties of the background interference. In this method, the running FFT of the detector output is evaluated recursively and the target is detected with the use of a threshold technique that separates the significant components of the local target and clutter spectra. In the illustrations, the motion of the target is used to generate a high frequency response at the output of each detector element etched with a mask that matches the point spread of the optical system.

1. Introduction

We consider the problem of detecting a target in the presence of strong interference. Unlike the usual methods the proposed approach is based on the design of a filter whose parameters are not specified in advance in terms of global statistics but are adaptively controlled in terms of local spectra evaluated in real time.

The problem is essentially multi-dimensional (space-time). However, for notational simplicity, we discuss only its one-dimensional form (time). The results can, in principle, be extended to several variables.

The one-dimensional problem in its post-detection form involves the estimation of a signal $s(t)$, or, at least, the determination of the presence of such a signal, in terms of the detector output

$$x(t) = c(t) + s(t) + \nu(t) \tag{1}$$

where $c(t)$ is the detector output due to clutter, and $\nu(t)$ is background noise. The processing is carried out digitally in terms of the samples

$$x[n] = x(nT)$$

of $x(t)$. Thus, the signal processing problem is the detection of the component $s[n]$ of the sum

$$x[n] = c[n] + s[n] + \nu[n] \tag{2}$$

---

* Work sponsored by DARPA Contract No. N00014-76C

† Polytechnic Institute of New York, Farmingdale, N.Y. 11735

The factors affecting the selection of the sampling interval T will not be considered.

The most common form of target detection uses a FIR filter whose output is the weighted sum

$$y[n] = \sum_{k=0}^{N-1} a_k x[n-k] \qquad (3)$$

The coefficients of this filter are independent of n and are chosen so as to yield a suitable frequency response

$$H(e^{j\omega T}) = \sum_{k=0}^{N-1} a_k e^{jk\omega T}$$

A special case is the mth difference filter obtained with $a_k = \binom{m}{k}$. The resulting system function is given by

$$H(z) = (1-z^{-1})^m$$

and can be realized as a cascade of first order systems. This filter is chosen primarily because it is simple (it requires no multiplication). It's frequency response is a rather primitive high-pass curve

$$\left| H(e^{j\omega T}) \right| = 2^m \left| \sin \frac{\omega T}{2} \right|^m$$

In the target detection problem it is desirable to adapt the system characteristics to the local properties of the background. This requires the design of a time-varying filter:

$$y[n] = \sum_{k=0}^{N-1} a_k[n]\, x\,[n-k] \qquad (4)$$

with adaptively controlled coefficients $a_k[n]$. The adaptation algorithms involve various numerical schemes for determining local statistics but are, in general, complex. A simple design, that can be used if the signal s[n] to be estimated is somehow available (as a pilot, for example, or as delayed observation), is the Widrow filter:

$$a_k[n] = a_k[n-1] + \mu \left(s[n]-y[n] \right) x\,[n-k] \qquad (5)$$

where μ is a suitable constant.

In the above filters, the processing is performed in the time domain. This is not optimum for the problem under consideration because the separation between target and clutter depends on frequency domain properties. FIR filters as in (3) or (4) can, of course, separate frequency components but this requires proper adjustment of all their coefficients a_k. The proposed processing involves processing directly in the frequency domain. As we shall see, the elimination of various frequency components is accomplished simply by eliminating the corresponding coefficients. This reduces drastically the number of the adaptively controlled parameters.

2. Running Spectra

The running FFT of a signal x[n] is by definition the sum

$$X_m[n] = \sum_{k=-M}^{M} x[n-k]w^{km} \qquad w=e^{j2\pi/N} \qquad N=2M+1 \tag{6}$$

Thus, $X_m[n]$ is the mth FFT coefficient of N consecutive samples of x[n] centered at n. The proposed adaptive frequency domain filter is a time-varying system whose output is the sum

$$z[n] = \sum_{m=-M}^{M} b_m[n]\ X_m[n] \tag{7}$$

where the weights $b_m[n]$ are adaptively controlled in a variety of ways depending on the applications. For example, if the Widrow algorithm is used, then $b_m[n]$ is determined as a first order recursion as in (5):

$$b_m[n] = b_m[n-1] + \mu\Big(s[n] - z[n]\Big)\ X_m[n] \tag{8}$$

It might appear that (7) is equivalent to (4), obtained merely by a linear transformation of the data. This, however, is not so. If it is concluded, either from prior information or from recent observations, that the frequency components of the interference are concentrated in certain frequency bands, the corresponding terms in (7) can be eliminated . This leads to the response

$$z[n] = 2\,\underline{\mathrm{Re}} \sum_{m=M_1[n]}^{M_2[n]} b_m[n]\ X_m[n] \tag{9}$$

where not only the coefficients $b_m[n]$ but also the cut-off frequencies $M_1[n]$ and $M_2[n]$ are adaptively controlled.

In the last section, we estimate the presence of s[n] in terms of the sum

$$z[n] = \frac{2}{N} \sum_{m=M_1[n]}^{M_2[n]} \underline{\mathrm{Re}}\ X_m[n] \tag{10}$$

This is a special case of (9) obtained with $b_m = 1/N$ and it equals s[n] if the frequencies M_1, M_2 separate completely the spectra of the signal and the interference. The determination of M_1 and M_2 is accomplished with a threshold method that is based on the determination of local clutter averages.

The advantages of the proposed filter are obvious: Processing in the frequency domain based not on global prior statistics but on local averages. However, it appears that, in contrast to time-domain filtering, the required number of arithmetic operations is large: N multiplications are required to determine $X_m[n]$ for each m and n. We shall presently show that this is not so. Each FFT $X_m[n]$ can be determined recursively with only one multiplication. Indeed, from (6) it follows that

$$X_m[n] - w^m X_m[n-1] = x[n+M]w^{-Mm} - x[n-M-1]w^{Mm} \tag{11}$$

that is $X_m[n]$ can be obtained as the output of a simple first order filter. To realize (11) in real time, we must of course introduce a delay of M units.

3. The Gemini Concept

Frequency domain filtering can be used in most methods of target detection because the suppression of the interference is based on the assumption that the clutter component c[n] of the detector output x[n] varies slowly relative to the target component s[n]. However, to be concrete, we shall consider a special case based on the Gemini principle (Fig.1):

Each detector element is covered with a mask consisting of vertical strips with transparency m(x) that is somehow matched to the point spread

$$h(x, y) = h(\sqrt{x^2 + y^2})$$

of the optical system and its output x(t) equals the integral of the light intensity across its surface. For simplicity, we assume that the center of the detector is at the origin (x = 0, y = 0) and that the target is a point source. The results can be readily generalized to arbitrary moving targets. Denoting by v_x, v_y the velocity components of the target properly scaled, we conclude that its image is $h(x-v_x t, y-v_y t)$. Hence, the de-

tector output is given by the integral

$$s(t) = \iint_D h(x-v_y t,\ y-v_y t)\, m(x)\, dxdy \tag{12}$$

where D is the region of the detector element. With

$$p(x) = \int_{-\infty}^{\infty} h(x, y)\, dy \tag{13}$$

the line-spread of the system, we obtain from (12), neglecting end-effects

$$s(t) = \int_{-\infty}^{\infty} p(x-v_y t)\, m(x)\, dx = \phi(v_x t) \tag{14}$$

where

$$\phi(x) = \int_{-\infty}^{\infty} p(x-\xi)\, m(\xi)\, d\xi \tag{15}$$

Denoting by H(u,v) the MTF of the system and by P(u), Φ(u), and M(u), the Fourier transforms of p(x), ϕ(x), and m(x) respectively, we obtain

$$P(u) = H(u, 0) \quad , \quad \Phi(u) = P(u)\, M(u) \tag{16}$$

The spectrum S(ω) of the detector output s(t) is thus given by

$$S(\omega) = \frac{1}{|v_x|}\ \Phi\left(\frac{\omega}{v_x}\right) = \frac{1}{|v_x|}\ H\left(\frac{\omega}{v_x}, 0\right) M\left(\frac{\omega}{v_x}\right) \tag{17}$$

This shows that a high velocity component v_x in the x-direction generates high frequencies in the component s(t) of x(t) due to the target. Hence, c(t) can be removed with frequency domain processing. The y-component of the velocity has no effect on the spectrum of x(t).

In the next section, we illustrate the above with a numerical example involving a one-dimensional mask as in Fig. 1. It might, however, be of interest to comment briefly on the possibility of detecting targets moving in any direction. As we show next, this can be done with masks consisting of circles:

$$m(x,y) = m(r) \tag{18}$$

that are matched somehow to the point-spread h(r)

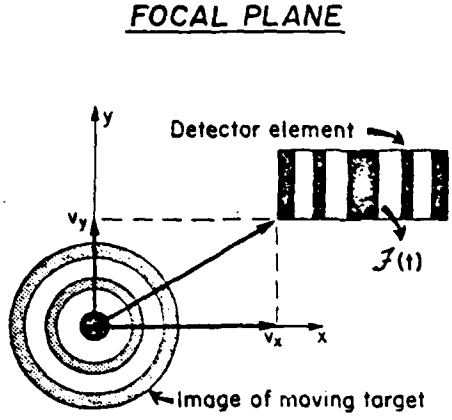


FIGURE 1 FOCAL PLANE IMAGE OF A MOVING TARGET.

We change the coordinates to (ξ, η) where ξ is in the direction of motion of the target. With v its velocity and η_o the distance from the origin to the line of motion, the image at time t is $h(\xi - vt, \eta-\eta_o)$ and the detector output is the integral

$$s(t) = \iint_D h(\xi - vt, \eta-\eta_o)\, m(\xi,\eta)\, d\xi d\eta \tag{19}$$

with

$$\phi(x,y) = \phi(r) = \iint h(x-\xi, y-\eta)\, m(\xi,\eta)\, d\xi d\eta = \phi(\sqrt{x^2+y^2}) \tag{20}$$

(19) yields

$$s(t) = \phi(vt, \eta_o) \tag{21}$$

Thus, the detector output x(t) is the profile $\phi(\xi,\eta_o)$ of $\phi(r)$ on the plane $\eta = \eta_o$ properly scaled. This curve is shown in Fig. 2 as a function

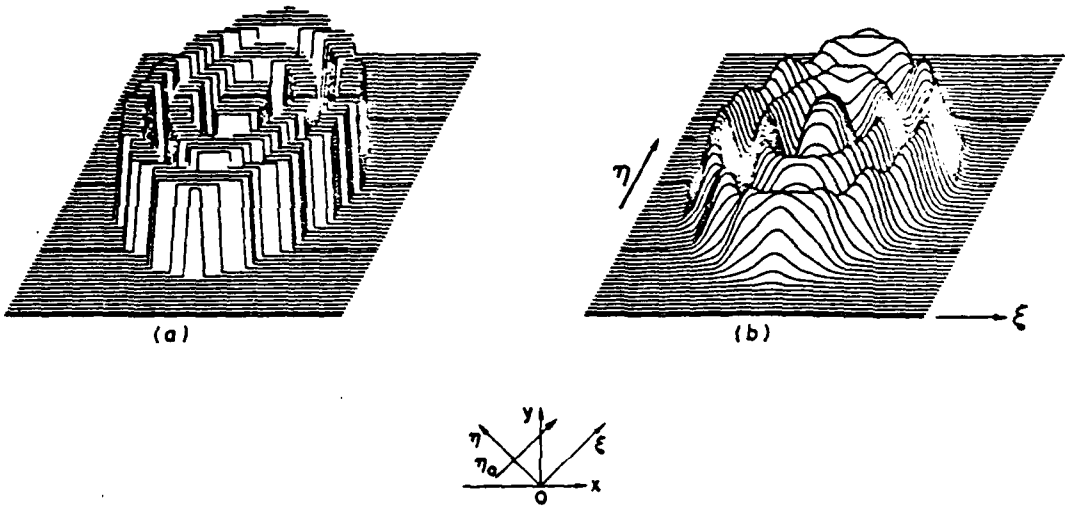


FIGURE 2 (a) CIRCULAR MASK
(b) DETECTOR OUTPUT $\phi(\xi,\eta)$ DUE TO A MOVING POINT SOURCE.

of ξ for various values of η_o. The point spread used is the Airy pattern

$$h(r) = \frac{J_1^2(r)}{r^2}$$

and the mask m(r) is a succession of transparent and opaque rings with boundaries at the zeros of $J_1(r)$.

4. Numerical results

In this section, we illustrate the adaptive frequency domain method with an example involving the detection of a moving target in the presence of strong interference. The data are computer generated.

The samples of the detector output form a discrete signal x[n] as in (2).

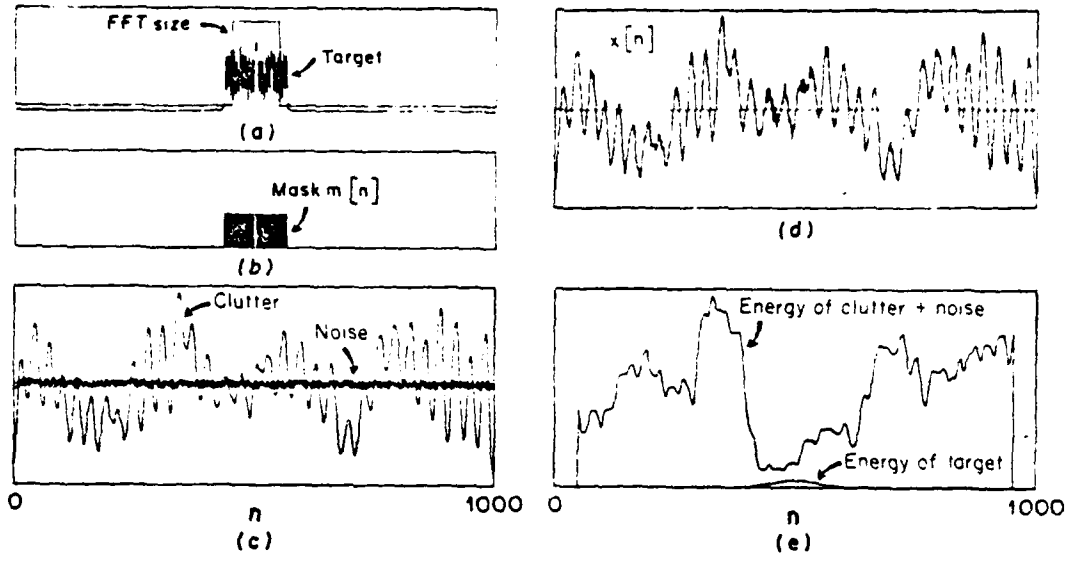


FIGURE 3 (a) TARGET s[n] AND SIZE OF THE RUNNING FFT. (b) m[n] : ETCHED MASK. (c) CLUTTER c[n] AND BACKGROUND NOISE ν[n]. (d) DETECTOR'S OUTPUT x[n]. (e) ENERGY OF s[n] AND c[n] + ν[n], AVERAGED OVER THE FFT SIZE.

(Fig.3) and our objective is to detect its presence. The numerical processing follows:

We form the running FFT $X_m[n]$ of x[n] of order

$$N = 101$$

using the first order recursion (11) and form the sum z[n] as in (10). The cut-off frequencies $M_1[n]$ and $M_2[n]$ are determined adaptively. The upper cut-off point depends on the noise component ν[n]. For simplicity we choose a fixed value $M_2[n] = 45$, limiting the discussion to the choice of $M_1[n]$. For this purpose, we form the intermediate average (Fig. 4)

$$\bar{X}_m[n] = \sum_{k=0}^{\infty} X_m[n-k]\, \alpha^k \qquad \alpha = 0.99 \tag{22}$$

of $X_m[n]$ and we choose for $M_1[n]$ the smallest value of M_1 such that

$$\left|\bar{X}_m[n]\right| < L \quad \text{for } M_1[n] - s \leq m < M_1[n] \tag{23}$$

where L is a threshold level

In figure 5a we show $\left|\bar{X}_m[n]\right|$ as a function of m and n, with $\left|\bar{X}_m[n]\right|$ truncated to the threshold level L. In figure 5b we show $\left|\bar{X}_m[n_o]\right|$ for $n_o = 700$ and in figure 5c we plot the values of the lower cut-off point $M_1[n]$ as a function of n.

The resulting sum

$$z[n] = \frac{2}{N} \sum_{m=M_1}^{M_2} \underline{Re}\ X_m[n]$$

is due primarily to the target but it contains a component e[n] (error) due to the frequency components of c[n] and $\nu[n]$ in the band (M_1, M_2).

We next form the <u>short term</u> and <u>long term</u> averages (Fig.6)

$$\overline{y[n]} = \sum_{k=0}^{\infty} y[n-k]\ \alpha_1^k \qquad \alpha_1 = 0.9$$

$$\overline{\overline{y[n]}} = \sum_{k=0}^{\infty} y[n-k]\ \alpha_2^k \qquad \alpha_2 = 0.999$$

of the energy $y[n] = z^2[n]$

These sums are determined recursively:

$$\overline{y[n]} = \alpha_1\ \overline{y[n-1]} + y[n] \qquad \overline{\overline{y[n]}} = \alpha_2\ \overline{\overline{y[n-1]}} + y[n]$$

Since the target is of short duration, the long-term average $\overline{\overline{y[n]}}$ is due namely to the clutter. If

$$\overline{y[n]} > k\ \overline{\overline{y[n]}} \qquad k=3$$

then the target is present. (Fig.7)

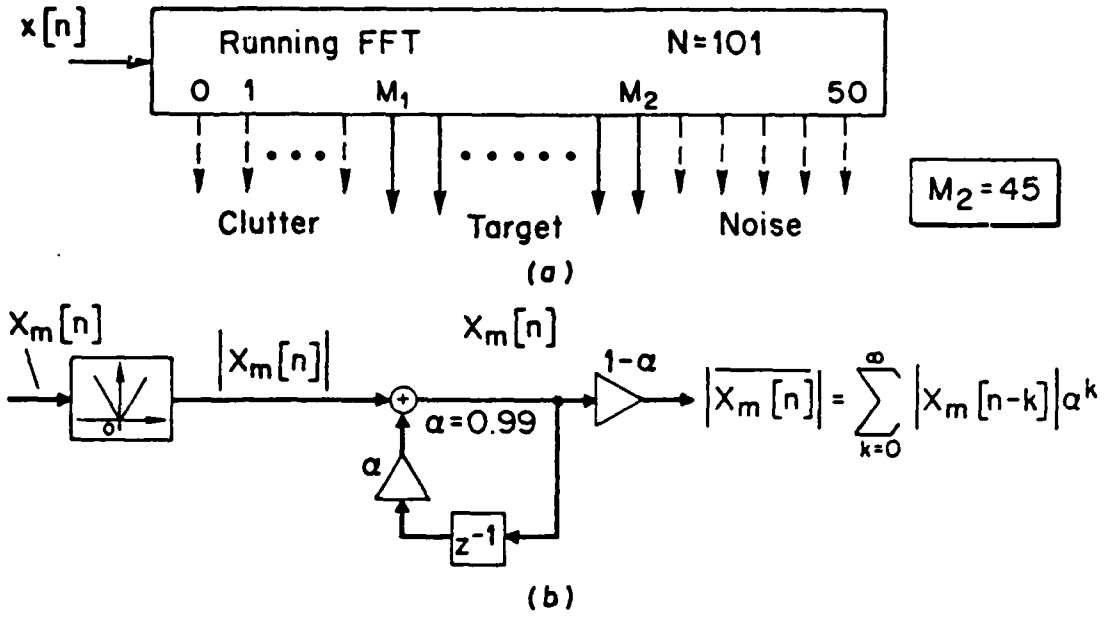


FIGURE 4 (a) RUNNING FFT
(b) GENERATION OF THE INTERMEDIATE AVERAGE $\bar{X}_m[n]$.

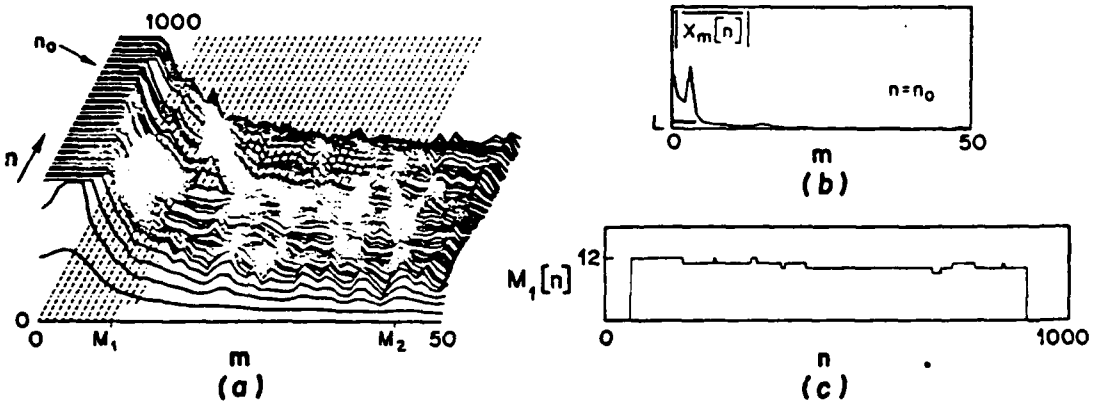


FIGURE 5 (a) $|\bar{X}_m[n]|$: INTERMEDIATE AVERAGE OF THE FREQUENCY COMPONENTS. (n is in steps of 20).

(b) $|\bar{X}_m[n]|$ FOR $n=n_0$

(c) $M_1[n]$: LOWER CUT OFF POINT OF THE RUNNING FILTER

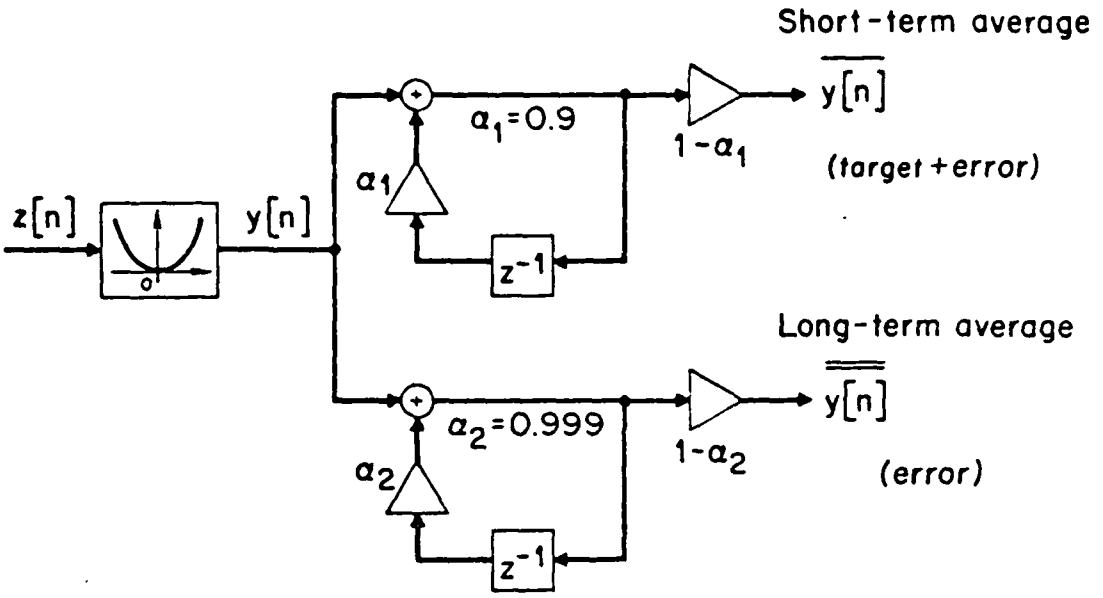


FIGURE 6 GENERATION OF THE SHORT-TERM, $\overline{y[n]}$, AND LONG-TERM, $\overline{\overline{y[n]}}$, AVERAGES.

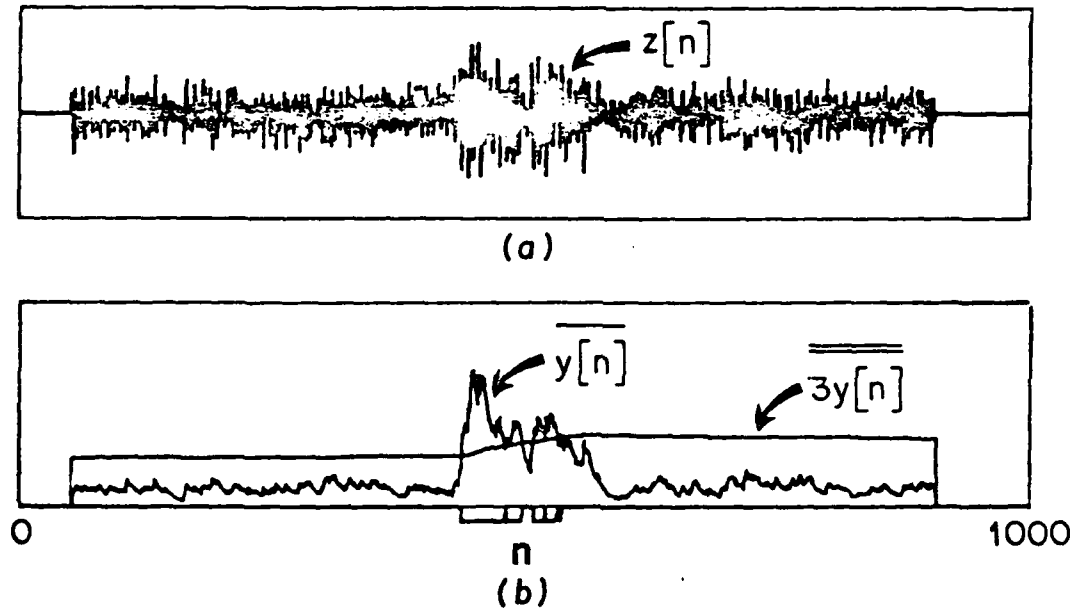


FIGURE 7 (a) z[n]: OUTPUT OF THE ADAPTIVE FILTER
(b) COMPARISON OF $\overline{y[n]}$ and $\overline{\overline{3y[n]}}$

13.

2. Bandlimited Extrapolation.

The investigation of the problems of extrapolating bandlimited signals by iteration was completed. The method was applied to problems in Image Enhancement, Spectral Estimation, Deconvolution, and Detection of Hidden Periodicities among others. The latest results are shown below:

"Windows and Extrapolation"
IEEE Workshop on Spectral Estimation, Cyprus Gardens, Florida, 1980.

"Detection of Hidden Periodicities by Adaptive Extrapolation"
IEEE Tr-ASSP-27, No. 5, October 1979 pp. 492-500.

Detection of Hidden Periodicities by Adaptive Extrapolation

ATHANASIOS PAPOULIS, *FELLOW, IEEE, AND* CHRISTODOULOS CHAMZAS

Abstract—A method is presented for determining the harmonic components of a noisy signal by nonlinear extrapolation beyond the data interval. The method is based on an algorithm that adaptively reduces the spectral components due to noise.

I. Introduction

An important problem in many applications is the determination of the frequency components of a signal

$$f(t) = \sum_{i=1}^{m} c_i e^{j\omega_i t} \tag{1}$$

in terms of the segment (data)

$$w_1(t) = \begin{cases} f(t) + n(t) & |t| < T \\ 0 & |t| > T \end{cases} \tag{2}$$

of $f(t)$ containing the noise component $n(t)$. The data are known for $|t| < T$ only for a variety of reasons:

1) The signal $f(t)$ can be written as a sum of exponentials for a limited time only (voice; nonstationary processes).

2) The available time of observation is limited (sun spots; weather trends).

3) Measurements are limited by instrument constraints (Michelson interferometer; diffraction-limited imaging).

Manuscript received November 22, 1978; revised January 25, 1979 and March 20, 1979. This work was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract N00014-76C 0144. This paper is in part from a Ph.D. dissertation submitted by C. Chamzas to the Faculty of the Polytechnic Institute of New York, Farmingdale, NY.

The authors are with the Department of Electrical Engineering, Polytechnic Institute of New York, Farmingdale, NY 11735.

0096-3518/79/1000-0492$00.75

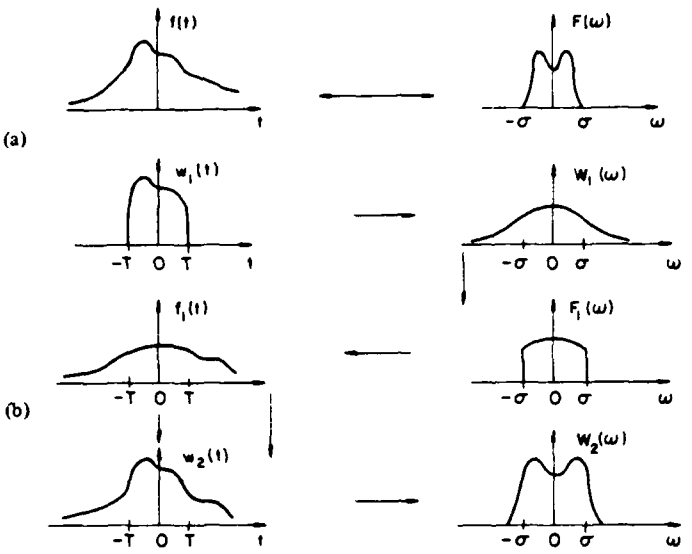


Fig. 1. (a) The unknown signal $f(t)$ and its Fourier transform $F(\omega)$. (b) First iteration starting with known segment $w_1(t)$.

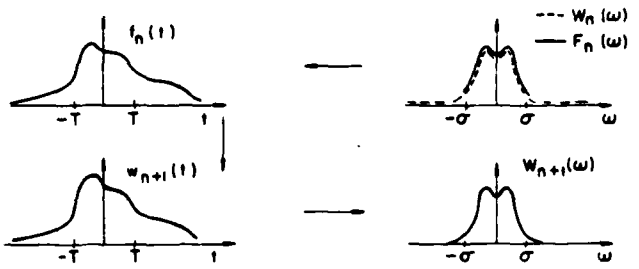


Fig. 2. nth iteration.

The unknown frequencies ω_i and coefficients c_i can be determined simply with ordinary Fourier transforms if the time of observation $2T$ is large compared to all the periods $T_i = 2\pi/\omega_i$ and their differences. This is not, however, the case if T is of the order $T_i - T_j$, particularly if the noise component $n(t)$ is not negligible. In this paper we present a method which, as we hope to show, is reliable even if T is small and the data are noisy.

The method involves only FFT and it is based on earlier results dealing with the problem of extrapolating band-limited functions [1], [2]. We review (for easy reference) the relevant parts of these results.

II. Extrapolation of Band-Limited Functions

Consider a function $f(t)$ with the Fourier transform $F(\omega)$ such that

$$F(\omega) = 0 \quad |\omega| > \sigma. \tag{3}$$

We form the function

$$w_1(t) = \begin{cases} f(t) & |t| < T \\ 0 & |t| > T \end{cases} \tag{4}$$

obtained by truncating $f(t)$ as in Fig. 1. We shall determine $f(t)$ in terms of $w_1(t)$ by numerical iteration.

First Step: We compute the Fourier transform $W_1(\omega)$ of $w_1(t)$ and form the function

$$F_1(\omega) = \begin{cases} W_1(\omega) & |\omega| < \sigma \\ 0 & |\omega| > \sigma. \end{cases} \tag{5}$$

We compute the inverse transform $f_1(t)$ of $F_1(\omega)$, and form the function

$$w_2(t) = \begin{cases} w_1(t) = f(t) & |t| < T \\ f_1(t) & |t| > T \end{cases} \tag{6}$$

and its Fourier transform $W_2(\omega)$.

This completes the First Step of the iteration (Fig. 1).

nth Step: We form the function (Fig. 2)

$$F_n(\omega) = \begin{cases} W_n(\omega) & |\omega| < \sigma \\ 0 & |\omega| > \sigma \end{cases} \tag{7}$$

where $W_n(\omega)$ is the function obtained at the end of the preceding step and compute the inverse transform $f_n(t)$ of $F_n(\omega)$. We form the function

$$w_{n+1}(t) = \begin{cases} f(t) & |t| < T \\ f_n(t) & |t| > T \end{cases} \tag{8}$$

and compute its Fourier transform $W_{n+1}(\omega)$.

If $f(t)$ is approximated by $f_n(t)$, the resulting mean-square error is given by

$$E_n = \int_{-\infty}^{\infty} [f(t) - f_n(t)]^2 \, dt = \frac{1}{2\pi} \int_{-\sigma}^{\sigma} |F(\omega) - F_n(\omega)|^2 \, d\omega. \tag{9}$$

We maintain that this error decreases twice at each iteration step. Indeed,

$$E_n = \int_{|t| < T} [f(t) - f_n(t)]^2 \, dt + \int_{|t| > T} [f(t) - f_n(t)]^2 \, dt.$$

But [see (7) and (8)]

$$\int_{|t| > T} [f(t) - f_n(t)]^2 \, dt = \int_{-\infty}^{\infty} [f(t) - w_{n+1}(t)]^2 \, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega) - W_{n+1}(\omega)|^2 \, d\omega$$

$$= \frac{1}{2\pi} \int_{|\omega| > \sigma} |F(\omega) - W_{n+1}(\omega)|^2 \, d\omega$$

$$+ \frac{1}{2\pi} \int_{-\sigma}^{\sigma} |F(\omega) - F_{n+1}(\omega)|^2 \, d\omega.$$

And since the last integral E_{n+1} [see (9)], we obtain

$$E_n - E_{n+1} = \int_{|t| < T} [f(t) - f_n(t)]^2 \, dt + \frac{1}{2\pi} \int_{|\omega| > \sigma} |W_{n+1}(\omega)|^2 \, d\omega \tag{10}$$

because $F(\omega) = 0$ for $|\omega| > \sigma$.

In [1] and [2] we show that $f_n(t) \to f(t)$ as $n \to \infty$. This is not true if the given segment $w_1(t)$ of $f(t)$ is noisy as in (2). In this case, a satisfactory estimate of $f(t)$ can be found by early termination of the iteration [2].

Note: From (10) it follows that the mean-square error E_n is a monoton decreasing function and since it is positive it tends to a limit. This does not prove the convergence of (9) because the limit need not be zero. It shows, however, that

$$E_n - E_{n+1} \to 0 \qquad n \to \infty.$$

Hence,

$$\int_{|t| < T} [f(t) - f_n(t)]^2 \, dt \to 0 \qquad n \to \infty. \tag{11}$$

Although the functions $f(t)$ and $f_n(t)$ are band limited, (11) *does not imply* that $f(t) \to f_n(t)$ because there is no lower bound on the energy concentration of band-limited functions in a finite interval [1], [3]. For example, the prolate spheroidal functions $\varphi_n(t)$ are band limited; their energy equals one but their energy concentration in the interval $(-T, T)$ tends to zero as $n \to \infty$. This is the case because the eigenvalues λ_n of the underlying integral equation tend to zero as $n \to \infty$.

We mention without elaboration that, in the discrete version of the problem, the convergence of the iteration can be deduced from (11) under suitable conditions. The reason is that the corresponding eigenvalues are finitely many, therefore, they have a positive minimum [4].

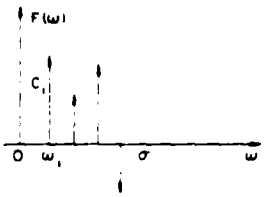


Fig. 3. Fourier transform of the unknown signal.

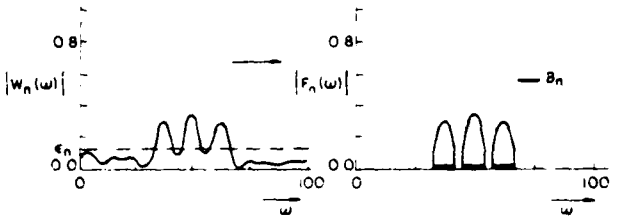


Fig. 4. Truncation of $W_n(\omega)$ below a threshold level ϵ_n yielding $F_n(\omega)$.

III. Adaptive Extrapolation

The preceding method was based on the assumption that the unknown function $f(t)$ is band limited. This information was used to reduce the error in the estimation of $f(t)$ twice at each iteration step. The speed of iteration can be increased and the effects of noise can be reduced if additional *a priori* information about $f(t)$ is available. Suppose, for example, that the size of the band of $F(\omega)$ is known but its precise location is unknown. We then choose a constant σ such that $F(\omega)$ vanishes outside the integral $(-\sigma, \sigma)$ and proceed as in Section II. As the iteration progresses, the form of $W_n(\omega)$ suggests appropriate reduction of the assumed band of $f(t)$.

The adaptive extrapolation method is particularly effective if $f(t)$ is a sum of exponentials as in (1). In this case, $F(\omega)$ consists of impulses (lines) as in Fig. 3:

$$F(\omega) = 2\pi \sum_{i=1}^{m} c_i \delta(\omega - \omega_i), \tag{12}$$

and our problem is to determine their locations ω_i and amplitudes c_i in terms of the known segment $w_1(t)$ of $f(t)$.

To solve this problem, we select a constant σ larger than the largest possible value of ω_i and we proceed with the iteration until $W_n(\omega)$ takes significant values only in a subset B_n of the band $(-\sigma, \sigma)$ of $f(t)$ (Fig. 4). This suggests that the unknown frequencies are in B_n. When this is observed, the function $F_n(\omega)$ of the nth iteration step is obtained from the following modification of (7)

$$F_n(\omega) = \begin{cases} W_n(\omega) & \omega \in B_n \\ 0 & \omega \in \bar{B}_n \end{cases} \tag{13}$$

(Fig. 4) where $\bar{B}_n$ is the complement of B_n. The process is repeated with occasional reduction of the size of B_n as further evidence suggests, and it terminates when $w_n(t)$ is essentially a sum of exponentials. Another application of the method is discussed in [5] in the context of deconvolution.

Discussion

The adaptive extrapolation method is essentially empirical. Although, as we see in the following examples, it works well in a number of cases, there is no *a priori* certainty that in a given problem it will converge to the unknown signal. In fact, if some of the components c_i of $f(t)$ are relatively small, they might be lost.

The accuracy and reliability of the method depends on a number of parameters: total number of unknown frequencies, possibly prior knowledge of this number, relative sizes of amplitudes c_i and frequencies ω_i, noise level, length $2T$ of the data interval, and available FFT size N. A precise statement, even empirical, of the importance of all these factors cannot be made; it would depend on many parameters. We are in the process of determining, empirically, the limits of the method for a number of special cases. We comment below, briefly, on certain empirical criteria for selecting the set B_n and on the limitations due to sampling.

For the subset B_n introduced in (13) we select the set of points such that the magnitude of $W_n(\omega)$ exceeds a threshold level ϵ_n:

$$\begin{aligned} |W_n(\omega)| &\geqslant \epsilon_n \qquad \omega \in B_n \\ |W_n(\omega)| &< \epsilon_n \qquad \omega \in \overline{B}_n. \end{aligned} \tag{14}$$

The choice of ϵ_n is dictated by two conflicting requirements: for a speedy convergence and noise reduction, ϵ_n must be large; it must be sufficiently small so that all frequency components of $f(t)$ remain in B_n. In the examples given below we used the following method for determining ϵ_n.

We first find the minimum M_{n-1} of $|W_{n-1}(\omega)|$ in the set B_{n-1}:

$$M_{n-1} = \min |W_{n-1}(\omega)| \qquad \omega \in B_{n-1}. \tag{15}$$

If ϵ_{n-1} is greater than μM_{n-1}, where μ is a constant less than one, then we do not change the threshold level. If ϵ_{n-1} is less than μM_{n-1} then we choose $\epsilon_n = \mu M_{n-1}$. Thus,

$$\epsilon_n = \max \{\epsilon_{n-1}, \mu M_{n-1}\}. \tag{16}$$

In the examples, μ is chosen between 0.9 and 0.99.

Numerical Considerations

The numerical implementation of the method involves the discrete signals

$$f_n = f(nt_o) \qquad F_n = F(n\omega_o)$$

obtained by sampling $f(t)$ and $F(\omega)$.

Suppose, first, that the problem is inherently discrete, i.e., that we wish to find the spectrum of a sequence f_n from incomplete data. Clearly, the discrete version of the iteration and of the band-limited assumption are self-evident. However, the assumption that $f(t)$ is a sum of sine waves has no obvious discrete version. It corresponds, loosely, to the assumption that the smallest distance of the nonzero frequencies is large compared to one (no "neighboring frequencies" are present). If this is the case, then the unknown frequencies can be determined exactly, provided that the data interval is not too small and the noise level is reasonable.

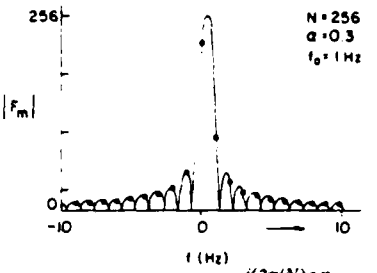


Fig. 5. Discrete spectrum $|F_m|$ of $f_{n_n} = e^{j(2\pi/N)\alpha n}$ ($\alpha = 0.3$, $N = 256$).

We turn now to our main problem: the numerical determination of the frequencies of an analog signal. We assume that the FFT size N is specified. It suffices, therefore, to select the size t_o of the sampling interval. As we know [1], the frequency interval is then determined because $\omega_o = 2\pi/Nt_o$. Since the data interval is $2T$, the number M of available samples equals $2T/t_o$. The choice of M is guided by the following considerations: if $M \ll N$, then the iteration might converge to the wrong frequencies. If M is large, then the aliasing errors are large.

It appears from our experience that $M = N/4$ is a reasonable compromise and it leads to $t_o \simeq 8T/N$. However, as we shall see, to increase the resolution we might use a larger value for t_o.

The accuracy of the method and the attainable resolution depend on the relationship between the unknown frequencies ω_i and the sampling frequency ω_o. If all unknown frequencies are multiples of ω_o

$$\omega_i = r_i \omega_o$$

then the problem is essentially discrete. If the unknown frequencies and their differences are large compared to ω_o, then the error is small because it is of the order of ω_o.

The problem of determining ω_i is difficult if ω_o is of the order of ω_i, and ω_i is not an integer multiple of ω_o

$$\omega_i = (r_i + \alpha)\, \omega_o \qquad |\alpha| < \tfrac{1}{2}.$$

In this case, the resolution error $\omega_o/2$ is of the order of ω_i. Furthermore, aliasing generates spurious frequencies in the vicinity of ω_i. Indeed, if

$$f(t) = e^{j\omega_i t}$$

then

$$f_n = e^{j\omega_i n t_o} = w^{(r_i+\alpha)n} \qquad w = e^{j2\pi/N}$$

yielding the discrete spectrum (Fig. 5)

$$F_m = \sum_{n=0}^{N-1} f_n w^{-mn} = \frac{1 - w^{(m - r_i - \alpha)N}}{1 - w^{(m - r_i - \alpha)}}.$$

To improve the accuracy, we can repeat the process with a larger value of t_o, using as starting B_o the set containing only the estimated frequencies ω_i and their neighbors.

IV. Illustrations

We illustrate the method with several examples involving signals whose unknown frequencies cannot be determined

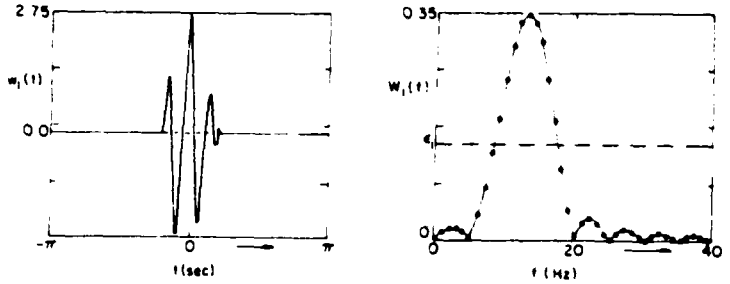


Fig. 6. Given segment $w_1(t)$ and its Fourier transform $W_1(\omega)$.

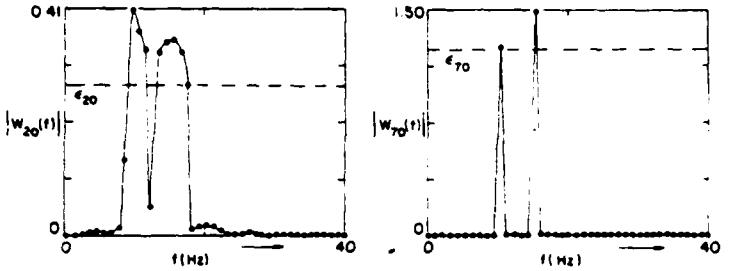


Fig. 7. Result of the iteration for $n = 20$ and $n = 70$.

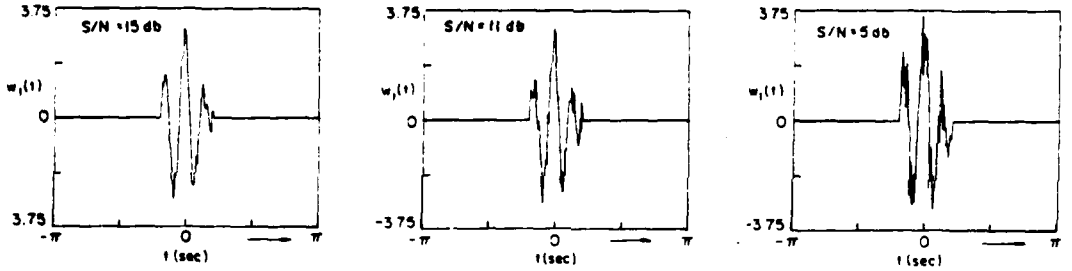


Fig. 8. Given data segment for $S/N = 15$, 11, and 5 dB.

from direct Fourier analysis. In these illustrations we consider several noise levels. With

$$w_1(t) = f(t) + n(t)$$

the given data, we define the signal-to-noise ratio S/N as the ratio of the energies of $f(t)$ and $n(t)$ in the data interval. In all examples, the noise is white and is uniformly distributed in the interval $(-c$ to $c)$. The ratio S/N is changed by changing the size of c.

The computations are carried out with

$$N = 256 \quad f_o = 1 \text{ Hz} \quad t_o = 1/256 \text{ s}.$$

To avoid large scaling factors, we divided all frequency components by $N/2$. In the examples we show also the value of the parameter μ [see (16)] and of the initial threshold level ϵ_1.

Example 1: The unknown signal is a sum of two sine waves

$$f(t) = 1.5 \cos(30\pi t + 60^\circ) + 1.25 \cos(20\pi t + 30^\circ)$$

and the unknown frequencies $f_1 = 10$ Hz and $f_2 = 15$ Hz are integral multiples of the sampling frequency ω_o.

a) We first assume that the data interval contains $M = 51$ sampling points and $n(t) = 0$.

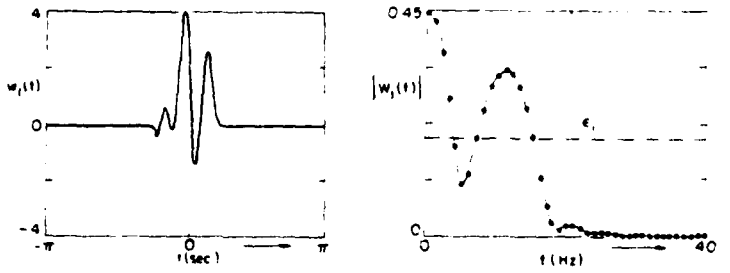
In Fig. 6 we show the given segment of the unknown signal and its spectrum. As we see from the figure, the frequencies f_1 and f_2 are not visible. The initial threshold is $\epsilon_1 = 0.15$ and its value at the nth iteration is obtained from (16) with $\mu = 0.99$. In Fig. 7 we show the results of the iteration for $n = 20$ and $n = 70$. At the 70th iteration the frequencies, amplitudes, and phases of $f(t)$ are recovered exactly.

We note that, in this case, the values of ϵ_1 and μ are not critical. Any value of μ between 0.9 and 0.99 and of ϵ_1 between 0.05 and 0.15 is adequate. The iteration was performed also with a data interval containing $M = 41$ sampling points. In this case, the results are similar but the speed of convergence is slower.

b) We consider, next, noisy data with various S/N ratios as in Fig. 8. In all cases,

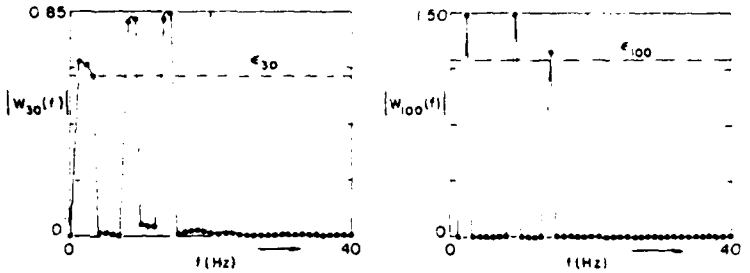
$$M = 51 \quad \mu = 0.99 \quad \epsilon_1 = 0.15.$$

The iteration was performed several times with the same signal but with different samples of noise. As the following indicates, the results are not the same for all samples: $S/N = 15$ dB ($c = 0.375$). Six samples were tried. In five of these, the frequencies f_1 and f_2 were found exactly. $S/N = 11$ dB ($c = 0.625$). Fourteen samples were tried. In nine, we ob-



(a) (b)

Fig. 9. (a) Given segment $w_1(t)$ of $f(t)$. (b) Fourier transform of $w_1(t)$.



(a) (b)

Fig. 10. Result of the iteration for $n = 30$ and $n = 100$.

tained f_1 and f_2 exactly. In four cases, an error of 1 Hz developed. In one case, the iteration yielded not two but three frequencies: $f_1 = 9$ Hz, $f_2 = 14$ Hz, and $f_3 = 15$ Hz. $S/N = 5$ dB ($c = 1.25$). This is a very noisy case. Of the eleven samples tried, three gave the correct answer, two yielded 1 Hz error, five resulted in 2 Hz error, and in one case the frequency $f_2 = 15$ Hz was lost.

Example 2: In this example $f(t)$ consists of three sine waves and the data are noiseless. We consider two cases. In the first case, the unknown frequencies are multiples of ω_o. In the second case, they are not.

a)

$$f(t) = 1.5 \cos 4\pi t + 1.5 \cos (18\pi t + 60^\circ) + 1.25 \cos (28\pi t + 30^\circ).$$

We start with the following values of the relevant parameters:

$$M = 59 \quad \mu = 0.95 \quad \epsilon_1 = 0.20.$$

In Fig. 9 we plot the given segment $f(t)$ and its spectrum. Fig. 10 shows the results of the iteration for $n = 30$ and $n = 100$. At the 100th iteration the frequencies, amplitudes, and phases of $f(t)$ are recovered exactly.

Again the values of μ and ϵ_1 are not critical. Essentially the same results are obtained if the data interval is reduced to $M = 51$ provided that μ is not less than 0.95.

The method has been tried also for a smaller data interval. However, the convergence is slow and the result inaccurate. With $M = 41$, $\mu = 0.99$, $\epsilon_1 = 0.20$ the component with the lowest frequency is lost.

b)

$$f(t) = 1.5 \cos 4.8\pi t + 1.5 \cos (18\pi t + 60^\circ) + 1.25 \cos (29.2t + 30^\circ).$$

In this case,

$$f_1 = (2 + 0.4) f_o \quad f_2 = 9 f_o \quad f_3 = (14 + 0.6) f_o.$$

We used $M = 59$, $\mu = 0.95$, and $\epsilon_1 = 0.20$.

With an FFT size $N = 256$, we obtained after 350 iteration steps the frequencies 2 Hz, 9 Hz, and 15 Hz (Fig. 11c).

Increasing the FFT size to $N = 512$, we found in 200 steps the frequencies 2.5 Hz, 8.75 Hz, 9.25 Hz, and 14.5 Hz. (Fig. 11d).

We note that the accuracy in the evaluation of coefficients of different levels can be improved if the threshold level ϵ_n is not constant through the band but it takes different values in the vicinity of each frequency. This is demonstrated in the next example.

Example 3: The unknown signal is a sum of five sine waves.

$$f(t) = 1.5 \cos 4\pi t + 1.25 \cos (12\pi t + 30^\circ) + 0.375 \cos (40\pi t + 60^\circ) + 0.625 \cos 50\pi t + 1.25 \cos (60\pi t + 45^\circ)$$

with frequencies 2, 6, 20, 25, and 30 Hz; the noise is zero. In Fig. 12 we show the given data, obtained with $M = 71$, and their spectrum. In the iteration we assume that $\mu = 0.99$ and $\epsilon_1 = 0.04$. The level of the threshold level at the nth iteration is defined as in (16). However, it is not constant throughout the band. Its value is determined from the behavior of $W_n(\omega)$ *in the vicinity of each maximum* (Fig. 13).

In Fig. 13 we show the iteration for $n = 10$ and $n = 20$. At the 50th step, (Fig. 14) we recover the frequencies 2, 6, 25, and 30. As it is clear from the figure, $W_n(\omega)$ contains a peak in the vicinity of $f = 20$. To determine its exact location we introduce the following variation to the method: we subtract from the given data the recovered portion of $f(t)$ and repeat the iteration starting with the new data $d(t)$ so obtained.

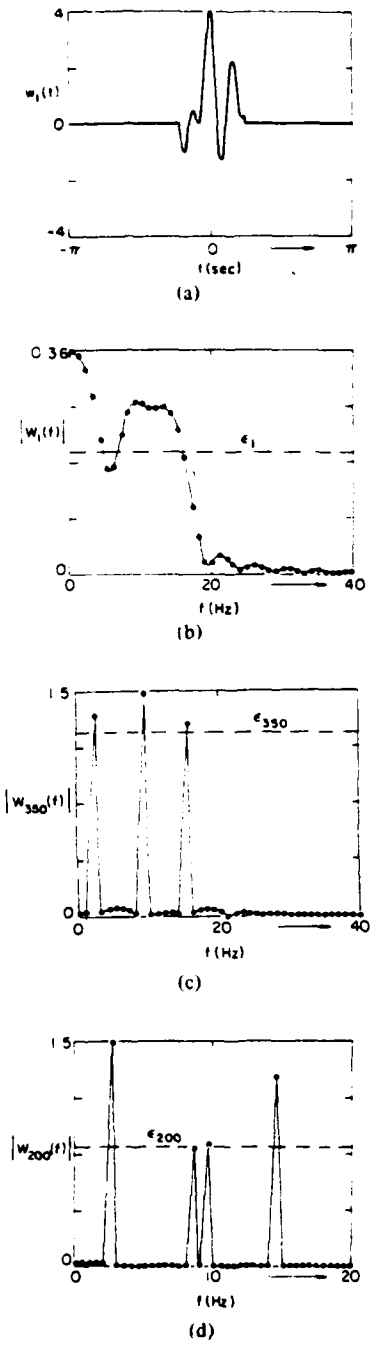


Fig. 11. (a) Given segment $w_1(t)$. (b) Fourier transform of $w_1(t)$. (c) Result of the iteration for $n = 350$ and FFT size $N = 256$. (d) Result of the iteration for $n = 200$ and FFT size $N = 512$.

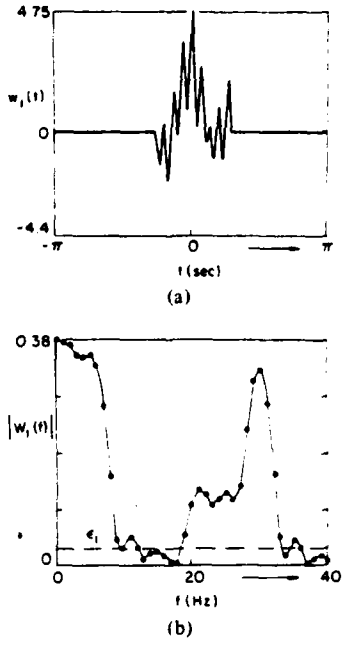


Fig. 12. (a) Given segment $w_1(t)$. (b) Fourier transform of $w_1(t)$.

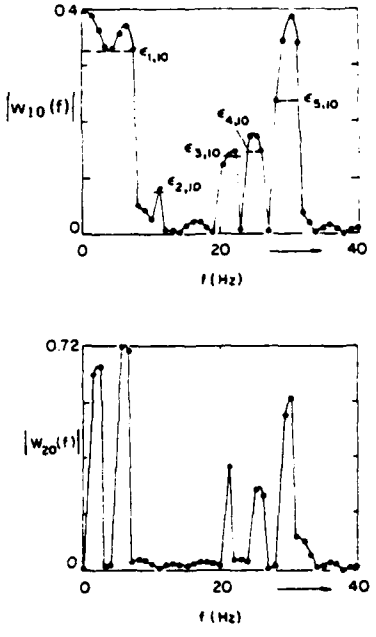


Fig. 13. Result of the iteration for $n = 10$ and $n = 20$.

In Fig. 15 we show $d(t)$ and its spectrum $D(\omega)$. The unknown frequency $f = 20$ is recovered at the 20th step (Fig. 16).

The iteration was performed also with a smaller data segment ($M = 61$). The results, however, were similar but the convergence slower.

Example 4: To test the limits of the method, we consider as a last case an example where the data interval is less than one-half the unknown period, and the unknown frequency is not a multiple of ω_o so that the aliasing is significant. We assume that

$$f(t) = 1.25 \cos (5.4\pi t + 30^\circ) \qquad T = 0.08 \text{ s}.$$

This yields $M = 41$ sampling points in the data interval.

The iteration was performed with $\mu = 0.99$ and $\epsilon_1 = 0.05$. We considered four different signal levels (Fig. 17).

a) $n(t) = 0$. At the 40th iteration we recover the frequency

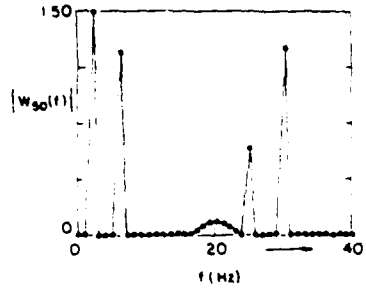
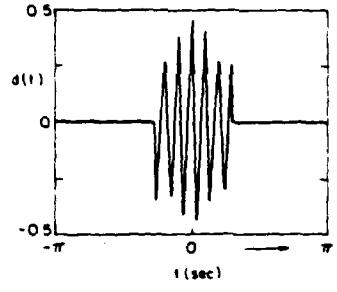
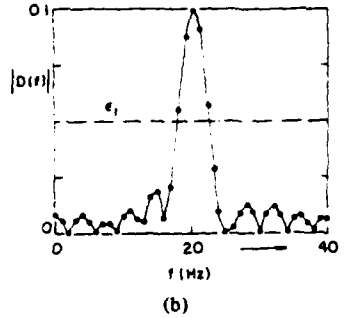


Fig. 14. Result of the iteration for $n = 50$.



(a)



(b)

Fig. 15. (a) New data segment $d(t)$. (b) Fourier transform of $d(t)$.

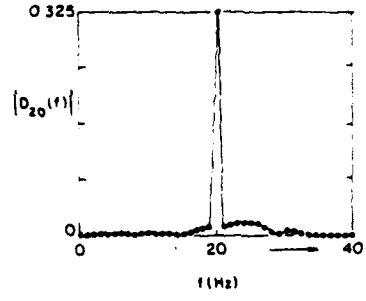
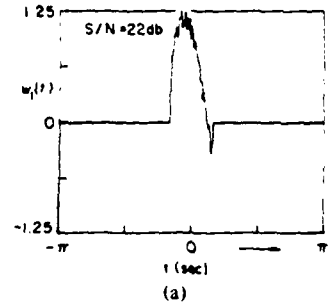
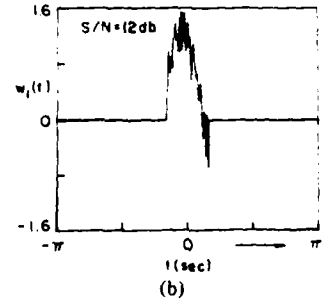


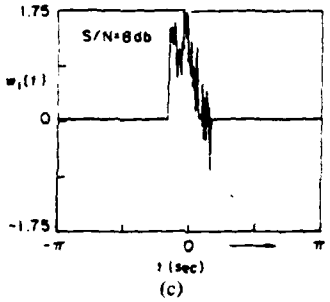
Fig. 16. Result of the iteration for $n = 20$.



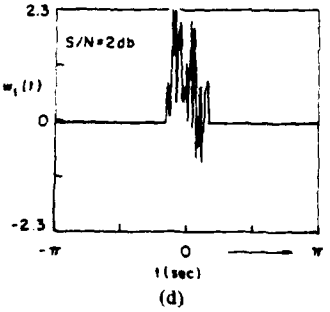
(a)



(b)



(c)



(d)

Fig. 17. Given data segment $w_1(t)$ for various noise levels. (a) S/N = 22 dB. (b) S/N = 12 dB. (c) S/N = 8 dB. (d) S/N = 2 dB.

$f = 3$ Hz. This is the nearest sampling frequency to the unknown $f_1 = 2.7$. However, since the resolution frequency $f_o = 1$ is of the order of f_1, the error is large. To reduce it, we increase the sampling interval from $t_o = 1/256$ to $t_o = 1/128$. This yields $f_o = 1/2$ Hz but the number of sampling points is reduced to $M = 21$. The iteration starts from the band B_o consisting of the location $f = 3$ of the recovered frequency and its two neighbors $f = 2.5$ and $f = 3.5$. After $n = 150$ steps, we recovered the frequency $f = 2.5$ (nearest to the unknown $f_1 = 2.7$).

The process was repeated with $t_o = 1/64$, that is, for $f_o = 1/4$ and $M = 11$ sampling points. The iteration yielded two frequencies: $f_a = 2.5$ and $f_b = 2.75$ with amplitudes $|c_a| = 0.616$, $|c_b| = 0.630$ and phases $\varphi_a = 30.06°$, $\varphi_b = 31.44°$, re-

TABLE I
RESULTS OF THE ITERATION FOR 20 NOISE SAMPLES

S/N	22 db	12 db	8 db	2 db
f_o (Hz)	(ϵ = 0.125)	(ϵ = 0.25)	(ϵ = 0.625)	(ϵ = 1.25)
	1(20)	1(18)	1(15)	1(12)
		2(2)	2(4)	2(4)
1			1(1)	1(1)
(M = 41)				3(1)
				4(3)
	2.5 (5)	2.5 (7)	2.5 (7)	
0.5	2.5, 3.0(8)	2.5, 3.0(4)	2.5, 3.0(1)	
(M = 21)	3.0(4)	3.0(9)	3.0(8)	
			3.5(2)	
			4.0(2)	
	2.75(7)	2.5 (9)		
0.25	2.5, 2.75(13)	2.5, 2.75(7)		
(M=11)		2.25, 2.5(2)		
		4.0(2)		

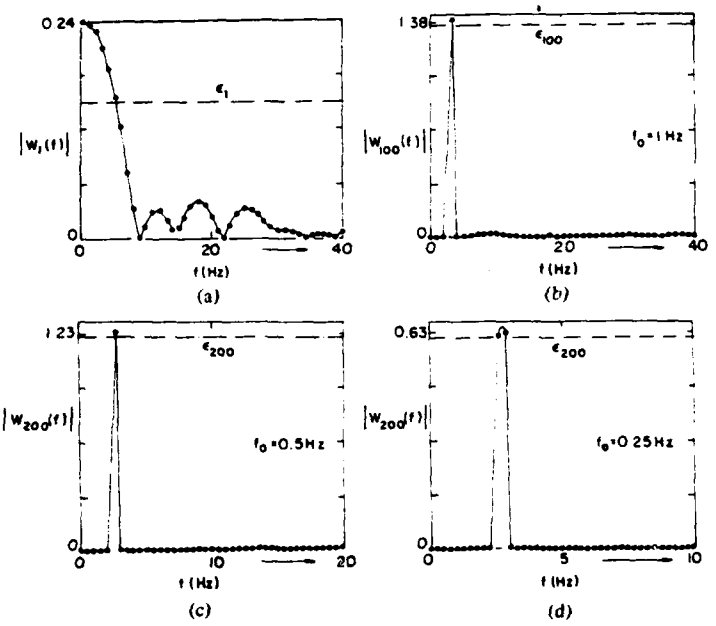


Fig. 18. $w_1(t) = 1.25 \cos(5.4\pi t + 30°) + n(t)$ $|t| < 0.08$ s. (a) Fourier transform of $w_1(t)$. (b) Result of the iteration for $f_o = 1$ Hz ($M = 41$) and $n = 100$. $f(t) = 1.38 \cos(6\pi t + 25.5°)$. (c) Result of the iteration for $f_o = 0.5$ Hz ($M = 21$) and $n = 200$. $f(t) = 1.23 \cos(5\pi t + 31.8°)$. (d) Result of the iteration for $f_o = 0.25$ Hz ($M = 11$) and $n = 200$. $f(t) = 0.62 \cos(5\pi t + 30.1°) + 0.63 \cos(5.5\pi t + 31.4°) \simeq 1.246 \cos(2.63\pi t + 30.7°)$.

spectively. The location $\hat{f}$ and amplitude $\hat{c}$ of the unknown frequencies was finally estimated by interpolation, yielding

$$\hat{f} = f_a + \frac{|c_b|}{|c_a| + |c_b|} f_o = 2.63 \text{ Hz}$$

$$\hat{c} = c_a + c_b = 1.246 \angle 30.76°.$$

b) $n(t) \neq 0$. We considered four different signal-to-noise ratios. All cases were performed 20 times using different samples for the noise. The statistical conclusions are described in Table I. The numbers in the table are the estimates in Hz of the unknown frequency. Numbers in parentheses indicate in how many samples out of 20 the cited estimates were obtained.

In Fig. 18 we show the results for $S/N = 8$ dB and $f_o = 1$, 0.5, 0.25 Hz.

REFERENCES

[1] A. Papoulis, *Signal Analysis*. New York: McGraw-Hill.
[2] —, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 735-742, Sept. 1975.
[3] D. Slepian, H. J. Landau, and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty principle I and II," *Bell Syst. Tech. J.*, vol. 40, no. 1, 1961.
[4] A. Papoulis and M. Bertran, "Digital filtering and prolate functions," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 674-681, Nov. 1972.
[5] A. Papoulis and C. Chamzas, "Improvement of range resolution by spectral extrapolation," *Ultrasonic Imaging*, vol. 1, pp. 121-135, Apr. 1979.

3. Undersampled Data.

The problem of estimating the spectrum $S(\omega)$ of a signal from undersampled data was considered. We showed that, although it is not possible in general to recover reliably $S(\omega)$, in special cases adequate estimates are possible. The results were presented in the following paper:

"Spectral Estimation from Random Samples"
IEEE International Conference on Information Sciences and Systems, Patras, Greece, 1979.

4. Spectral Estimation.

The fundamental problem of estimating the spectrum $S(\omega)$ of a random signal in terms of a single realization was considered with emphasis on the method of Maximum Entropy. Recent results led to the following two papers:

"Entropy: From first Principles to Spectral Estimation"
IEEE Tr-ASSP Workshop on Spectral Estimation, Hamilton, Ontario, 1981.

"Maximum Entropy and Spectral Estimation"
IEEE Tr-ASSP (to appear).

MAXIMUM ENTROPY AND SPECTRAL ESTIMATION: A REVIEW*

Athanasios Papoulis

Polytechnic Institute of New York
Route 110
Farmingdale, N.Y. 11743

ABSTRACT

The method of maximum entropy is reviewed with emphasis on its relationship to entropy rate, Wiener filters, autoregressive processes, extrapolation, the Levinson algorithm, lattice, all-pole, and all-pass filters and stability.

*This work was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract No. N00014-76-C-0144.

I. Introduction

In the last decade, several papers have been published discussing a method of spectral estimation based on the principle of maximum entropy [1]-[4] and the relationship of this method to entropy rate [5], the Wiener theory of prediction [6], [7] autoregressive processes, the Levinson algorithm [8], lattice filters [9], all-pole and all-pass filters, and stability. However, it appears that no single publication in the open literature explains simply the interconnection of these topics. The purpose of this paper is an attempt to do so starting from first principles [10]. The effectiveness of the method in the solution of specific problems will not be considered here. In the Appendix, we comment briefly on its conceptual justification. The material is developed with some originality; however, the paper is essentially tutorial.

The entire development is based on the orthogonality principle [11]: In the estimation of a random variable y by a linear combination

$$\hat{y} = a_1 x_1 + \cdots + a_N x_N \tag{1}$$

of the N random variables $x_1, \ldots, x_N$ (data), the MS error

$$P = E\{(y - \hat{y})^2\} \tag{2}$$

is minimum if the estimation error

$$e = y - \hat{y} \tag{3}$$

is orthogonal to the data x_k, that is, if

$$E\{ex_k\} = 0 \quad k = 1, \ldots, N \tag{4}$$

The resulting MS error P is then given by

$$P = E\{e^2\} = E\{ey\} \tag{5}$$

We state also for later use the following results from the theory of linear systems with stochastic inputs [11]: Suppose that the input to a discrete linear system is a stationary process $x[n]$ with autocorrelation

$$R_{xx}[m] = E\{x[n+m]\,x[n]\} \tag{6}$$

and power spectrum

$$S_{xx}(z) = \sum_{m=-\infty}^{\infty} R_{xx}[m]z^{-m} \tag{7}$$

If $h[n]$ is the delta response and $H(z)$ the system function of the system, then the power spectrum of the resulting output $y[n] = x[n]*h[n]$ is given by

$$S_{yy}(z) = S_{xx}(z)\,H(z)\,H(1/z) \tag{8}$$

In the above we assumed, as we shall throughout the paper, that all processes and systems are real. With trivial modifications, the results hold also for complex processes.¶ The spectral estimation problem has two parts:

1. Deterministic Estimate the power spectrum $S(z)$ of a process $s[n]$ in terms of the $N+1$ values $R[0], R[1], \ldots, R[N]$ of its autocorrelation.

2. Random Estimate the power spectrum $S(z)$ of a process $s[n]$ in terms of the N_o values $s[1], s[2], \ldots, s[N_o]$ of a single realization of $s[n]$.

As we show in the paper, the maximum entropy solution of Part 1 can be presented as a recursive modification of the Wiener prediction filter. The modification is based on the Levinson algorithm expressed in terms of forward and backward predictors. The solution of Part 2 is given by an estimator whose various parameters satisfy the same equations as in the deterministic case, with the only difference that in the evaluation of the recursion coefficient Γ_N [see (53)], all ensemble averages are replaced by suitable time-averages.

II. Prediction

We wish to estimate the future value $s[n+1]$ of a random signal $s[n]$ in terms of the sum

$$\hat{s}_N[n+1] = a_1^N s[n] + \ldots + a_N^N s[n-N+1] = \sum_{k=1}^{N} a_k^N s[n-k+1] \qquad (9)$$

involving its N most recent values $s[n-k]$. The set of weights a_k^N that minimizes the MS value of the prediction error defines a FIR filter of order N (Fig. 1) called the forward predictor (one-step) of $s[n]$. The superscript N in a_k^N specifies the order of the predictor. Since $s[n]$ is stationary, the optimum weights a_k^N are independent of n. We can give, therefore, to the variable n in (9) any value. With

$$\hat{e}_N[n] = s[n] - \hat{s}_N[n] \qquad (10)$$

the forward predictor error, we have [see (4)]

$$E\{\hat{e}_N[n]\, s[n-k]\} = 0 \qquad 1 \le k \le N \qquad (11)$$

This yields the system

$$\begin{aligned} R[0]a_1^N + R[1]a_2^N + \ldots + R[N-1]a_N^N &= R[1] \\ R[1]a_1^N + R[0]a_2^N + \ldots + R[N-2]a_N^N &= R[2] \\ \text{-----------------------------------} & \\ R[N-1]a_1^N + R[N-2]a_2^N + \ldots + R[0]a_N^N &= R[N] \end{aligned} \qquad (12)$$

expressing the predictor coefficients a_k^N in terms of the $N+1$ values $R[0], \ldots, R[N]$ of the autocorrelation $R[m]$ of $s[n]$. In the next section, we discuss a recursion method for solving this system.

Applying (5) to our estimator, we conclude that the MS estimation error $\hat{P}_N$ is given by

$$\hat{P}_N = E\{\hat{e}_N^2[n]\} = E\{\hat{e}_N[n]s[n]\} = R[0] - \sum_{k=1}^{N} a_k^N R[k] \qquad (13)$$

Note: Consider two processes $s[n]$ and $s_o[n]$ with autocorrelations $R[m]$ and $R_o[m]$, respectively. From (12) it follows that, if

$$R[m] = R_o[m] \quad \text{for} \quad 0 \leq m \leq N \qquad (14)$$

then the Nth order predictors of $s[n]$ and $s_o[n]$ are identical. Conversely, if the Nth order predictors of $s[n]$ and $s_o[n]$ are identical, then (12) shows that $R[m] = c\,R_o[m]$ $0 \leq m \leq N$. The proportionality constant c equals 1 if $R[m]$ and $R_o[m]$ satisfy one additional equation, for example, if the Nth order MS errors are equal or if $R_o[0] = R[0]$, that is, if the two processes have the same average power

$$P_0 = E\{s^2[n]\} = R[0] \qquad (15)$$

The prediction error [see (10)]

$$\hat{e}_N[n] = s[n] - \sum_{k=1}^{N} a_k^N s[n-k] \qquad (16)$$

is the output of a system (Fig. 1) with input $s[n]$ and system function

$$\hat{H}_N(z) = 1 - \sum_{k=1}^{N} a_k^N z^{-k} \qquad (17)$$

This system will be called the forward error filter.

The backward predictor. We shall now estimate the process $s[n]$ in terms of the backward predictor

$$\check{s}_N[n] = \sum_{k=1}^{N} b_k^N s[n+k]$$

involving its N closest future values $s[n+k]$. With

$$\check{e}_N[n] = s[n] - \check{s}_N[n] \qquad (18)$$

the backward predictor error, we have as in (11)

$$E\{\check{e}_N[n]s[n+k]\} = 0 \qquad 1 \le k \le N \tag{19}$$

This yields the system

$$\sum_{r=1}^{N} b_r^N R[r-k] = R[k] \qquad 1 \le k \le N$$

which is identical to the system (12). Hence, $b_k^N = a_k^N$, that is, the backward predictor of $s[n]$ is the sum

$$\check{s}_N[n] = a_1^N s[n+1] + \ldots + a_N^N s[n+N] \ . \tag{20}$$

The predictor error is thus given by

$$\check{e}_N[n] = s[n] - \sum_{k=1}^{N} a_k^N s[n+k] \tag{21}$$

Denoting by $\check{P}_N$ its MS value, we conclude as in (13) that

$$\check{P}_N = E\{\check{e}_N[n]s[n]\} = R[0] - \sum_{k=1}^{N} a_k^N R[k]$$

In other words, the forward and backward MS predictor errors are equal:

$$\check{P}_N = \hat{P}_N = P_N \tag{22}$$

Clearly, $\check{e}_N[n]$ is the output of a system with input $s[n]$ and system function

$$\check{H}_n(z) = 1 - \sum_{k=1}^{N} a_k^N z^k \tag{23}$$

This system will be called the backward error filter. Comparing with (17), we conclude that

$$\check{H}_N(z) = \hat{H}_N(1/z) \tag{24}$$

In the above, we assumed that $s[n]$ is a real process. The results hold also for complex processes subject to the following modifications

$$R[-m] = R^*[m] \qquad b_k^N = (a_k^N)^* \qquad \check{H}_N(z) = \hat{H}_N^*(1/z^*)$$

Autoregressive processes. An autoregressive process (AR) of order M is a random signal $s[n]$ satisfying the recursion equation

$$s[n] - c_1 s[n-1] - \cdots - c_M s[n-M] = \zeta[n] \tag{25}$$

where $\zeta[n]$ is stationary white noise with

$$R_{\zeta\zeta}[m] = \sigma^2 \delta[m] \qquad S_{\zeta\zeta}(z) = \sigma^2 \tag{26}$$

From the definition it follows that $s[n]$ is the output of a linear system with input $\zeta[n]$ and system function

$$T(z) = \frac{1}{1 - \sum_{k=1}^{M} c_k z^{-k}} \tag{27}$$

If this system is stable, then $s[n]$ is a stationary process given by

$$s[n] = \sum_{r=0}^{\infty} h[r]\, \zeta[n-r] \tag{28}$$

where $h[n]$ is the causal inverse transform [12] of $H(z)$. This shows that for any $k \geq 1$, the random variable $s[n-k]$ is a linear combination of only the past values of $\zeta[n]$, hence

$$E\{s[n-k]\, \zeta[n]\} = 0 \qquad k \geq 1 \tag{29}$$

because $\zeta[n]$ is white noise by assumption.

We maintain that the predictor $\hat{s}_N[n]$ of $s[n]$ of order $N \geq M$ is the sum

$$\hat{s}[n] = c_1 s[n-1] + \cdots + c_M s[n-M] \tag{30}$$

and the predictor error $\hat{e}_N[n] = s[n] - \hat{s}_N[n]$ equals the process $\zeta[n]$:

$$\hat{e}_N[n] = \zeta[n] \tag{31}$$

Indeed, as we see from (25), $s[n] - \hat{s}_N[n] = \zeta[n]$. Hence, the orthogonality condition (11) follows from (29). The resulting MS error is given by

$$P_N = E\{\hat{e}_N^2[n]\} = E\{\zeta^2[n]\} = \sigma^2$$

We note that the sum in (30) is of the form (9) where

$$a_k^N = \begin{cases} c_k & 1 \le k \le M \\ 0 & M < k \le N \end{cases} \tag{32}$$

With $\hat{H}_N(z)$ the error filter as defined in (17), it follows from (27) and (32) that the system function $T(z)$ of the system (AR filter) specified by (25) is given by

$$T(z) = \frac{1}{\hat{H}_N(z)} \tag{33}$$

Clearly, $s[n]$ is the output of $T(z)$ with input $\zeta[n]$. Denoting by $S(z)$ its power spectrum, we conclude from (8) and (33) that

$$S(z) = \frac{S_{\zeta\zeta}(z)}{\hat{H}_N(z)\hat{H}_N(1/z)} \tag{34}$$

Since $S_{\zeta\zeta}(z) = \sigma^2$, the above yields

$$\bar{S}(\omega) \equiv S(e^{j\omega T}) = \frac{\sigma^2}{\left|1 - \sum_{k=1}^{M} c_k e^{-jk\omega T}\right|^2} \tag{35}$$

Extrapolation. The inverse transform of $S(z)$ is the autocorrelation $R[m]$ of $s[n]$. From (12) and (32) it follows that

$$R[N] = \sum_{k=1}^{M} c_k R[N-k] \quad \text{for every } N \geq M \tag{36}$$

Setting $N = M$, we obtain a system of M equations expressing the M coefficients c_k in terms of the first $M+1$ values of $R[m]$. With c_k so determined, we can use (36) to evaluate successively $R[N]$ for every $N > M$. Thus, (36) is an extrapolation formula for $R[N]$.

III. The Levinson Algorithm

The solution of the system (12) involves the inversion of the matrix

$$\begin{bmatrix} R[0] & R[1] & \cdots & R[N-1] \\ R[1] & R[0] & \cdots & R[N-2] \\ \cdots & \cdots & \cdots & \cdots \\ R[N-1] & R[N-2] & \cdots & R[0] \end{bmatrix} \tag{37}$$

This matrix has a special form (Toeplitz [13]) and can be inverted easily by a simple iteration known as Levinson's algorithm [8]. We shall present the result as a recursion involving directly the predictor coefficients.

Theorem. The forward predictor $\hat{s}_N[n]$ can be written as a sum

$$\hat{s}_N[n] = \hat{s}_{N-1}[n] + \Gamma_N\left(s[n-N] - \check{s}_{N-1}[n-N]\right) \tag{38}$$

where

$$\begin{aligned} \hat{s}_{N-1}[n] &= \sum_{k=1}^{N-1} a_k^{N-1} s[n-k] \\ \check{s}_{N-1}[n-N] &= \sum_{k=1}^{N-1} a_k^{N-1} s[n-N+k] \end{aligned} \tag{39}$$

are the forward and backward predictors of $s[n]$ and $s[n-N]$ respectively, and the coefficient Γ_N is a constant to be determined. Equation (38) can be expressed in terms of the forward and backward predictor errors

$$\hat{e}_{N-1}[n] = s[n] - \hat{s}_{N-1}[n] \qquad \check{e}_{N-1}[n-N] = s[n-N] - \check{s}_{N-1}[n-N]$$

Indeed, subtracting both sides from $s[n]$, we obtain

$$\hat{e}_N[n] = \hat{e}_{N-1}[n] - \Gamma_N \check{e}_{N-1}[n-N] \tag{40}$$

It suffices, therefore, to prove (40).

Proof by induction. Clearly $\hat{s}_N[n]$ is a linear combination of the N most recent values $s[n-k]$ of the signal. It suffices, therefore, to show that $\hat{e}_N$ satisfies the orthogonality condition. By the induction hypothesis, we assume that the sequences $\hat{e}_{N-1}$ and $\check{e}_{N-1}$ are the predictor errors of order N - 1, that is, [see (11) and (19)]

$$\begin{aligned} E\{\hat{e}_{N-1}[n]\, s[n-k]\} &= 0 \qquad 1 \le k \le N-1 \\ E\{\check{e}_{N-1}[n-N]\, s[n-k]\} &= 0 \qquad 1 \le k \le N-1 \end{aligned} \tag{41}$$

We shall show that if $\hat{e}_N$ is given by (40), then it is the Nth order predictor error. As we know, this is true if

$$E\{\hat{e}_N[n]\, s[n-k]\} = 0 \qquad 1 \le k \le N \tag{42}$$

From (41) it follows that (42) holds for $1 \le k \le N-1$. It suffices, therefore, to select Γ_N such as to satisfy (42) for $k=N$: $E\{\hat{e}_N[n]\, s[n-N]\} = 0$. Inserting (40) into the above, we obtain

$$E\{\hat{e}_{N-1}[n]\, s[n-N]\} = \Gamma_N E\{\check{e}_{N-1}[n-N]\, s[n-N]\} \tag{43}$$

and since [see (39)]

$$E\{\hat{e}_{N-1}[n]\, s[n-N]\} = E\left\{\left(s[n] - \sum_{k=1}^{N-1} a_k^{N-1} s[n-k]\right) s[n-N]\right\}$$

$$= R[N] - \sum_{k=1}^{N-1} a_k^{N-1} R[N-k]$$

and

$$E\{\check{e}_{N-1}[n-N]\,s[n-N]\} = P_{N-1}$$

(43) yields

$$P_{N-1}\Gamma_N = R[N] - \sum_{k=1}^{N-1} a_k^{N-1} R[N-k] \tag{44}$$

We have, thus, expressed Γ_N in terms of the coefficients a_k^{N-1} of the predictor of order N - 1 and the corresponding MS error P_{N-1}. This error is given also by [see (22)]

$$P_{N-1} = E\{\hat{e}_{N-1}[n]\,s[n]\} \tag{45}$$

With this choice of Γ_N, (42) holds for every k from 1 to N.

Using (38), we can express a_k^N in terms of a_k^{N-1} and the constant Γ_N. Indeed, with $\hat{s}_N[n]$ as in (9), we obtain equating coefficients of both sides of (38) the recursive equation

$$a_k^N = a_k^{N-1} - \Gamma_N\, a_{N-1}^{N-1} \quad 1 \le k \le N-1 \quad a_N^N = \Gamma_N \tag{46}$$

where Γ_N is determined from (44). Since this equation involves the MS error P_{N-1}, to complete the induction, we must determine its Nth order value

$$P_N = E\{\hat{e}_N[n]\,s[n]\} \tag{47}$$

We maintain that

$$P_N = (1 - \Gamma_N^2)\,P_{N-1} \tag{48}$$

To show this, we insert $\hat{e}_N[n]$, as given by (40), into (47) and use (45) and the fact that

$$E\{\check{e}_{N-1}[n-N]\,s[n]\} = E\left\{\left(s[n-N] - \sum_{k=1}^{N-1} a_k^{N-1}\, s[n-N+k]\right) s[n]\right\}$$
$$= R[N] - \sum_{k=1}^{N-1} a_k^{N-1} R[N-k] = \Gamma_N P_{N-1}$$

The result is Equation (48). The induction starts with $\Gamma_0 = 0$, $P_0 = E\{s^2[n]\} = R[0]$ and for $N = 1$ it yields $P_0\Gamma_1 = R[1]$, $P_1 = P_0(1 - \Gamma_1^2)$.

The recursion (38) and its equivalent (40) hold also for the backward predictors. Reasoning similarly, we obtain

$$\check{s}_N[n-N] = \check{s}_{N-1}[n-N] + \Gamma_N\left(s[n] - \hat{s}_{N-1}[n]\right) \tag{49}$$

$$\check{e}_N[n-N] = \check{e}_{N-1}[n-N] - \Gamma_N \hat{e}_{N-1}[n] \tag{50}$$

Lattice. Equations (40) and (50) can be given the following graphical interpretation [9], [14]. In Fig. 2, we show N lattice sections connected in cascade. Each section consists of one delay element and two multipliers. The input to the system so formed equals $s[n] = \hat{e}_o[n] = \check{e}_o[n]$ and the two outputs equal the forward and backward predictor errors.

We note that the transfer functions from the input A to the two outputs B and C equal $\hat{H}_N(z)$ and

$$z^{-N}\check{H}_N(z) = z^{-N}\hat{H}_N(1/z)$$

respectively, where $\hat{H}_N(z)$ is the forward error filter and $\check{H}_N(z)$ is the backward error filter.

Note: We derive next for later use a modified form of (44). Clearly, [see (41)],

$$P_N = E\{e_N^2[n]\} = E\left\{\left(\hat{e}_{N-1}[n] - \Gamma_N \check{e}_{N-1}[n-N]\right)^2\right\} \tag{51}$$

Since the coefficients of the predictor minimize P_N, we must have

$$\frac{\partial P_N}{\partial \Gamma_N} = 0 = E\left\{-2\left(\hat{e}_{N-1}[n] - \Gamma_N \check{e}_{N-1}[n-N]\right)\check{e}_{N-1}[n-N]\right\}$$

Hence,

$$P_{N-1}\Gamma_N = E\{\hat{e}_{N-1}[n]\check{e}_{N-1}[n-N]\} \tag{52}$$

The above can be written in the symmetrical form

$$\Gamma_N = \frac{E\{\hat{e}_{N-1}[n]\check{e}'_{N-1}[n-N]\}}{\frac{1}{2}E\{\hat{e}^2_{N-1}[n] + \check{e}^2_{N-1}[n-N]\}} \tag{53}$$

This is a consequence of the fact that the forward and backward MS errors are equal. It can also be derived by writing P_N in the symmetrical form

$$P_N = \frac{1}{2}E\{\hat{e}^2_N[n] + \check{e}^2_N[n-N]\} = \frac{1}{2}E\Big\{\Big(\hat{e}_{N-1}[n] - \Gamma_N \check{e}_{N-1}[n-N]\Big)^2 +$$

$$+ \Big(\check{e}_{N-1}[n-N] - \Gamma_N \hat{e}_{N-1}[n]\Big)^2\Big\} \tag{54}$$

and minimizing with respect to Γ_N.

Stability. We have shown that the Nth order MS error P_N is given by $P_N = (1 - \Gamma_N^2)P_{N-1}$. Since $P_N \leq P_{N-1}$, it follows that

$$|\Gamma_N| \leq 1 \tag{55}$$

with equality iff $P_N = P_{N-1}$. We shall use this result to show that the forward error filter

$$\hat{H}_N(z) = 1 - \sum_{k=1}^{N} a_k^N z^{-k} \tag{56}$$

is a Hurwitz polynomial, i.e., all its roots z_i are inside the unit circle [14]:

$$|z_i| \leq 1 \tag{57}$$

From this it will follow that all roots of the backward error filter $\check{H}'_N(z)$ are outside the unit circle because

$$\check{H}'_N(z) = \hat{H}_N(1/z) \tag{58}$$

Proof by induction.* Clearly, $\hat{H}_1(z) = 1 - \Gamma_1 z^{-1}$, hence, $|z_1| = |\Gamma_1| \leq 1$. Suppose that (57) is true for all orders up to N-1. We shall show that it is

*This proof was suggested to the author by Th. Andrikos.

true for order N. From (40) and (58), it follows that

$$\hat{H}_N(z) = \hat{H}_{N-1}(z) - \Gamma_N z^{-N} \check{H}_{N-1}(z) = \hat{H}_{N-1}(z) - \Gamma_N z^{-N} \hat{H}_{N-1}(1/z) \quad (59)$$

If z_i is a root of $\hat{H}_{N-1}(z)$, then by the induction hypothesis, $|z_i| \leq 1$. And, since $1/z_i$ is a root of $\hat{H}_{N-1}(1/z)$, we conclude that the ratio

$$H_a(z) = \frac{z^{-N} \hat{H}_{N-1}(1/z)}{H_{N-1}(z)} \quad (60)$$

is an all-pass system, hence, it can be factored into a product

$$H_a(z) = \frac{z z_1^* - 1}{z - z_1} \cdots \frac{z z_{N-1}^* - 1}{z - z_{N-1}}$$

From this it follows that

$$|H_a(z)| \begin{cases} < 1 & |z| > 1 \\ = 1 & |z| = 1 \\ > 1 & |z| < 1 \end{cases} \quad (61)$$

because this holds for every bilinear term of $H_a(z)$. To complete the proof, we shall show that if z_o is a root of $\hat{H}_N(z)$, then $|z_o| \leq 1$. Setting $z = z_o$ in (59), we obtain

$$\hat{H}_N(z_o) = 0 = \hat{H}_{N-1}(z_o) - \Gamma_N z_o^{-N} \hat{H}_{N-1}(1/z_o)$$

Hence

$$\frac{1}{\Gamma_N} = \frac{z_o^{-N} \hat{H}_N(1/z_o)}{\hat{H}_{N-1}(z_o)} = H_a(z_o)$$

But $|\Gamma_N| \leq 1$, therefore, $|H_a(z_o)| > 1$ and (61) yields $|z_o| \leq 1$.

IV. Spectral Estimation

We shall now relate the preceding results to the method of maximum entropy.

Deterministic case. We are given the first N+1 values of the autocorrelation $R[m]$ of a random process $s[n]$ and we wish to estimate its power spectrum

$$S(z) = \sum_{m=-\infty}^{\infty} R[m]z^{-m} \tag{65}$$

For this purpose, we shall construct an AR process $s_o[n]$ of order N with autocorrelation $R_o[m]$ such that $R_o[m] = R[m]$ for $|m| \leq N$. The power spectrum $S_o(z)$ of this process will be used as the estimate of $S(z)$.

The construction of $s_o[n]$ is based on the determination of the Nth order prediction

$$\hat{s}_N[n] = \sum_{k=1}^{N} a_k^N s[n-k] \tag{66}$$

of $s[n]$. As we have shown, the coefficients a_k^N of this predictor can be determined by solving the system (12), or equivalently, from the recursion equations

$$\begin{aligned} a_k^N &= a_k^{N-1} + \Gamma_N a_{N-k}^{N-1} && 1 \leq k \leq N-1 \\ P_{N-1}\Gamma_N &= R[N] - \sum_{k=1}^{N} a_k^{N-1} R[N-k] && N \geq 1 \\ a_N^N &= \Gamma_N \qquad P_N = (1-\Gamma_N^2)P_{N-1} \end{aligned} \tag{67}$$

with the initial condition $P_0 = R[0]$. In either case, the solution is uniquely determined in terms of the known values of $R[m]$. We next form the AR filter of Fig. 3 with system function

$$T(z) = \frac{1}{\hat{H}_N(z)} \quad \text{where} \quad \hat{H}_N(z) = 1 - \sum_{k=1}^{N} a_k^N z^{-k} \tag{68}$$

is the error filter of the predictor $\hat{s}_N[n]$ of $s[n]$ [see (17)]. As input to this system we use a stationary white noise process $\zeta[n]$ with average power P_N. Denoting by $s_o[n]$ the resulting output, we conclude that

$$s_o[n] - \sum_{k=1}^{N} a_k^N s_o[n-k] = \zeta[n] \qquad R_{\zeta\zeta}[m] = P_N \delta[m] \tag{69}$$

The system T(z) is stable because $\hat{H}_N(z)$ is a Hurwitz polynomial. Therefore, its output $s_o[n]$ is stationary and since it satisfies (69), it is AR. From this and (32), it follows that the Nth order predictor $\hat{s}_o[n]$ of $s_o[n]$ is given by

$$\hat{s}_o[n] = \sum_{k=1}^{N} a_k^N s_o[n-k] \tag{70}$$

This shows that the process $s_o[n]$ of Fig. 3 and the original process $s[n]$ have identical predictors, therefore [see (14)], their corresponding autocorrelations $R_o[m]$ and $R[m]$ are equal for $|m| \leq N$ within a factor. We maintain that this factor is one. Indeed, with $\hat{e}_o[n] = s_o[n] - \hat{s}_o[n]$ the prediction error of $s_o[n]$, it follows from (69) and (70) that

$$E\{\hat{e}_o^2[n]\} = E\{\zeta^2[n]\} = R_{\zeta\zeta}[0] = P_N$$

hence (see note, page 4)

$$R_o[m] = R[m] \qquad |m| \leq N \tag{71}$$

This shows that if we use as the estimate of the unknown spectrum S(z) of $s[n]$ the spectrum $S_o(z)$ of the AR process $s_o[n]$, its inverse transform will agree with the given values of $R[m]$. Since $S_{\zeta\zeta}(z) = P_N$, if follows from (34) that

$$S_o(z) = \frac{P_N}{\hat{H}_N(z)\,\hat{H}_N(1/z)} \tag{72}$$

and on the unit circle,

$$\overline{S}_o(\omega) = S_o(e^{j\omega T}) = \frac{P_N}{\left|1 - \sum_{k=1}^{N} a_k^N e^{-jk\omega T}\right|^2} \tag{73}$$

This is the maximum entropy estimate of the unknown spectrum $\overline{S}(\omega)$. The numerator P_N and the coefficients a_k^N are determined from (67).

Random case. We are given the N_o samples (data) $s[1], s[2], \ldots, s[N_o]$ of a single realization of a process $s[n]$ and we wish to estimate its power spectrum $\overline{S}(\omega)$. The maximum entropy estimator $\hat{\overline{S}}(\omega)$ of $\overline{S}(\omega)$ is an all-pole function as in (73). However, unlike the deterministic case, the value of N is not specified. The problem now is to select first N and then to estimate the coefficients a_k^N. Suppose that we have somehow decided on the value of N. We then proceed as in the deterministic case using the recursion equations (67). These equations specify a_k^N in terms of the constants Γ_N and the initial condition $P_0 = R[0]$. It suffices, therefore, to determine the estimates $\hat{\Gamma}_N$ and $\hat{P}_0$ of these constants by appropriate time-averages involving the given data.

For the estimate of P_0, we use the sum

$$\hat{P}_0 = \frac{1}{N_o} \sum_{n=1}^{N} s^2[n] \tag{74}$$

For the estimate of Γ_N, we use the time-average version of (44) or (53). As we have shown, these equations are equivalent; however, because of end-effects the corresponding time-averages are not equivalent. We shall use the latter

because, unlike (44), it leads to an estimate $\hat{\Gamma}_N$ that satisfies the stability condition

$$|\hat{\Gamma}_N| \leq 1 \tag{75}$$

Our problem, thus, is reduced to the determination of the time-average form of the equation

$$\Gamma_N = \frac{E\{\hat{e}_{N-1}[n]\check{e}_{N-1}[n-N]\}}{\frac{1}{2}E\{\hat{e}^2_{N-1}[n] + \check{e}^2_{N-1}[n-N]\}} \tag{76}$$

where

$$\hat{e}_{N-1}[n] = s[n] - \left(a_1^N s[n-1] + \ldots + a_{N-1}^{N-1} s[n-N+1]\right)$$

$$\check{e}_{N-1}[n-N] = s[n-N] - \left(a_1^N s[n-N+1] + \ldots + a_{N-1}^{N-1} s[n-1]\right)$$

The above involves all samples of $s[n]$ from n to n-N. And since the data are available only from $n = 1$ to $n = N_o$, to avoid overflow in the time-average form of (76), we must limit the values of n from $N+1$ to N_o. This interval has $N_o - N - 1$ points, hence,

$$\hat{\Gamma}_N = \frac{\frac{1}{N_o - N - 1}\sum_{n=N+1}^{N_o} \hat{e}_{N-1}[n]\check{e}_{N-1}[n-N]}{\frac{1}{2(N_o - N - 1}\sum_{n=N+1}^{N_o}\left(\hat{e}^2_{N-1}[n] + \check{e}^2_{N-1}[n-N]\right)} \tag{77}$$

The above ratio satisfies (75) because (Schwarz inequality)

$$\left|\sum \hat{e}_{N-1}[n]\check{e}_{N-1}[n-N]\right|^2 \leq \sum \hat{e}^2_{N-1}[n] \sum \check{e}^2_{N-1}[n-N]$$

and

$$\sqrt{|xy|} \leq \tfrac{1}{2}(|x| + |y|)$$

With $\hat{\Gamma}_N$ so determined, the Nth order estimate of the unknown spectrum is given by

$$\hat{\bar{S}}_N(\omega) = \frac{\hat{P}_N}{\left|1 - \sum_{k=1}^{N} \hat{a}_k^N e^{-jk\omega T}\right|^2} \tag{78}$$

where the coefficients are determined recursively as in (67):

$$\hat{a}_k^N = \hat{a}_k^{N-1} + \hat{\Gamma}_N \hat{a}_{N-k}^{N-1} \qquad \hat{a}_N^N = \hat{\Gamma}_N \tag{79}$$

$$\hat{P}_N = (1 - \hat{\Gamma}_N)\, \hat{P}_{N-1} \tag{80}$$

The recursion starts with the estimate (76) of P_0.

We conclude with a brief comment on the choice of N. This choice is dictated by two conflicting requirements: For a satisfactory approximation of the unknown spectrum $\bar{S}(\omega)$ by an all-pole function $\bar{S}_o(\omega)$, N should be as large as possible. However, in the estimate (77) of $\hat{\Gamma}_N$, the number of terms in the time-average equals $N_o - N - 1$, and as we know, this number should be large for the variance of the estimate to be small. Various schemes have been suggested for selecting N but they will not be discussed [15], [16].

The estimate (79) of Γ_N can be obtained by minimizing the time-average form

$$I_N = \frac{1}{2(N_o - N - 1)} \sum_{m=N+1}^{N_o} \left(\hat{e}_{N-1}[n] - \Gamma_N \check{e}_{N-1}[n-N]\right)^2 + \left(\check{e}_{N-1}[n-N] - \Gamma_N \hat{e}_{N-1}[n]\right)^2$$

of the MS error P_N as given by (54). Setting $\partial I_N / \partial \Gamma_N = 0$, we obtain (77). However, the resulting value of I_N does not equal the estimate $\hat{P}_N$ of P_N obtained recursively from (80).

V. Maximum Entropy

We shall finally show that the all-pole model is a consequence of the principle of maximum entropy. The required background is discussed in the Appendix. We repeat the problem: We are given the N+1 values R[0], ..., R[N] of the autocorrelation R[m] of a random process s[n] and we wish to estimate its power spectrum $\overline{S}(\omega)$. The statistics of s[n] are determined in terms of the joint density of the r.v. s[n], s[n-1], ..., s[n-r]. Hence, to apply the method of maximum entropy, we must determine the unknown values of R[m] so as to maximize the entropy $H(s_0, \cdots, s_r)$ of these r.v. and to find the limit as $r \to \infty$. This is equivalent to the maximization of the entropy rate H_s of s[n] [see (A-12)], subject to the given constraints. We shall show that H_s is maximum if s[n] is a normal process with power spectrum as in (73).

We give three proofs. The first two involve the maximization of H_s. In the third, we find R[N+1] by maximizing $H(s_0, \ldots, s_{N+1})$, and, with R[N+1] so determined, we continue the process. This method can be questioned because it does not yield the maximum of $H(s_0, \ldots, s_N, \cdots, s_{N+k})$ subject to the given constraints. However, the result is correct in the limit as $k \to \infty$.

Method 1. We form the Nth order predictor $\hat{s}_N[n]$ of s[n] and the predictor error

$$s[n] - \sum_{k=1}^{N} a_k^N s[n-k] = \hat{e}_N[n] \tag{81}$$

Clearly, $\hat{e}_N[n]$ is the output of the error filter $\hat{H}_N(z)$ [see (17)] with input s[n]. Hence, [see (A-6)], its entropy rate $H_{\hat{e}}$ is given by

$$H_{\hat{e}} = H_s + \frac{1}{4\omega_o} \int_{-\omega_o}^{\omega_o} \ln |\hat{H}_N(e^{j\omega T})|^2 \, d\omega \tag{82}$$

To maximize H_s, it suffices, therefore, to maximize $H_{\hat{e}}$ because the integral is specified in terms of the given values of R[m]. As we know,

$$E\{e_N^2[n]\} = P_N = R[0] - \sum_{k=1}^{N} a_k^N R[k] \tag{83}$$

Therefore, $H_{\hat{e}}$ is maximum if the process $\hat{e}_N[n]$ is normal white noise (see Appendix). And since $\hat{e}_N[n]$ is the right side of the recursion equation (81), we conclude that the optimum $s[n]$ is an AR process of order N, hence, its power spectrum is all-pole as in (73).

We note that the optimum $s[n]$ is a normal process because it is the output of the stable linear system $T(z) = 1/\hat{H}_N(z)$ whose input is the normal process $\hat{e}_N[n]$.

Method 2. In the following reasoning, we assume that the process $s[n]$ is normal. As we have just shown, this assumption is not restrictive. From the normality of $s[n]$ it follows that, within a constant, its entropy rate is given by [see (A-20)]

$$H_s = \frac{1}{4\omega_o} \int_{-\omega_o}^{\omega_o} \ln \bar{S}(\omega)\, d\omega \qquad \omega_o = \frac{\pi}{T} \tag{84}$$

where

$$\bar{S}(\omega) = \sum_{m=-\infty}^{\infty} R[m] e^{-jm\omega T} \tag{85}$$

Since $R[m]$ is specified for $|m| \leq N$, the above integral depends on the values of $R[m]$ for $|m| > N$ and it is maximum if

$$\frac{\partial H_s}{\partial R[m]} = 0 = \frac{1}{4\omega_o} \int_{-\omega_o}^{\omega_o} \frac{1}{\bar{S}(\omega)} e^{-jm\omega T} d\omega \qquad |m| > N \tag{86}$$

This shows that the Fourier series coefficients of the function $1/\bar{S}(\omega)$ are zero for $|m| > N$. Hence,

$$\frac{1}{\overline{S}(\omega)} = \sum_{k=-N}^{N} c_k e^{-jk\omega T} \tag{87}$$

And since $\overline{S}(\omega) \geq 0$, it follows from the Fejér-Riess theorem [12] that the above sum can be written as a square. This yields

$$\overline{S}(\omega) = \frac{1}{\left| \sum_{k=0}^{N} b_k e^{-jk\omega T} \right|^2} \tag{88}$$

We have thus shown that $\overline{S}(\omega)$ is an all-pole function as in (73) where $P_N = 1/|b_o|^2$ and $a_k^N = b_k/b_o$.

Method 3. This method is iterative. In the first iteration, we determine $R[N+1]$ so as to maximize the entropy H of the r.v. $s[n]$, $s[n-1]$, ..., $s[n-N-1]$. For this purpose, we start with the assumption that $R[N+1]$ is specified and we determine the joint density of the above r.v. for maximum H. As we show in the Appendix, H is maximum if these r.v. are jointly normal with zero mean. In this case [see (A-24)]

$$H = \ell n \sqrt{(2\pi e)^{N+1} \Delta} \tag{89}$$

where

$$\Delta \equiv \begin{vmatrix} R[0] & R[1] \ldots R[N+1] \\ R[1] & R[0] \ldots R[N] \\ \ldots\ldots\ldots\ldots\ldots \\ R[N+1] & R[N] \ldots R[0] \end{vmatrix}$$

The above determinant is a non-negative quadratic in $R[N+1]$ and it is maximum if

$$R[N+1] = \sum_{k=1}^{N} a_k^N R[N+1-k] \tag{90}$$

where the coefficients a_k^N satisfy (12). With $R[N+1]$ so determined, we continue the iteration, and at the rth step we determine $R[N+r]$ so as to maximize the entropy of the r.v.

$$s[n], \ldots, s[n-N-r] \tag{91}$$

This yields the extrapolation formula

$$R[N+r] = \sum_{k=1}^{N} a_k^N R[N+r-k] \tag{92}$$

The coefficients a_k^{N+r} of the predictor of $s[n]$ of order $N+r$ satisfy again the system (12), where now N is replaced by $N+r$. From this and (92) it follows that

$$a_k^{N+r} = 0 \quad \text{for} \quad N < k \leq N+r$$

This shows that the Nth order predictor is also the predictor of any higher order; hence, $s[n]$ is an AR process of order N and its power spectrum is an all-pole function as in (88) [see also (35) and (36)].

APPENDIX

ENTROPY AND ENTROPY RATE

We present next for easy reference the relevant concepts from the theory of entropy.

Consider a probability space S and a partition A of S, that is, a countable collection of mutually exclusive events A_i whose union equals S.

Definition. The entropy H(A) of A is the sum

$$H(A) = -\sum_{i=1}^{N} p_i \ln p_i \quad \text{where} \quad p_i = P(A_i) \tag{A-1}$$

Thus, entropy is a number associated to each partition of a probability space. This number has the following significance. As we know, if the experiment is performed n times and the event A_i occurs n_i times, then "almost certainly"

$$p_i \simeq n_i/n \tag{A-2}$$

provided that n is "sufficiently large." This heuristic statement is the basis for the use of probability in real problem. It can be given a precise interpretation in the context of the law of large numbers [11].

We shall call each sequence of the forms t = {A_i occurs $n_i \cong np_i$ times in a specific order } typical. The union of all such sequences will be denoted by T. Clearly,

$$P(T) \simeq 1 \tag{A-3}$$

because according to (A-2), the typical sequences occur "almost certainly." Each typical sequence is an event in the product space $S^n = S \times \cdots \times S^n$ and

$$P(t) = p_1^{n_1} p_2^{n_2} \cdots p_N^{n_N} \tag{A-4}$$

Since $n_i \cong np_i$ and $p_i = e^{\ell n\, p_i}$, it follows from (A-1) that

$$P(t) \cong e^{np_1 \ell n p_1} \cdots e^{np_N \ell n\, p_N} = e^{-nH(A)} \qquad (A\text{-}5)$$

Hence, the total number N_T of typical sequences is given by [5]

$$N_T \simeq e^{nH(A)} \qquad (A\text{-}6)$$

It follows readily from (A-1) that $H(A) \leq \ell nN$ with equality iff $p_i = 1/N$. And since the total number of sequence in S^n equals N^n, we conclude that if all p_i's are not equal, then

$$H(A) < \ell nN \quad N_T << N^n$$

for large n. Thus, although $P(T) \simeq 1$, the number N_T of sequences in T is small compared with the total number N^n of all possible sequences. It is this result that forms the basis for the applications of entropy. We shall use it to establish the conceptual equivalence between maximum entropy and the classical definition of probability.

Suppose first that we wish to determine p_i in the absence of any prior information (no constraints). In this case, all sequences in S^n are equaly likely, hence, N_T must be nearly equal to N^n because $P(T) \simeq 1$. From this and (A-6), it follows that H(A) must equal its maximum ℓnN.

Suppose next that prior information is available in the form of inequality constraints, or expected values. Such information leads to the condition that only certain sequences in the space S^n are admissible, forming the subset S_c^n. All typical sequences are now in S_c^n, and since $P(T) \simeq 1$ and the sequences in S_c^n are equally likely, N_T must contain most of them, i.e., H(A) must be maximum subject to the given constraints.

The above argument is imprecise in the same sense as (A-2), however, as in that case, it can be given a precise interpretation as a limit theorem.

A consequence of the conceptual equivalence between maximum entropy and classical definition is the conclusion that the former is subject to the same critique as the latter. We should note in support of maximum entropy that in most problems involving prior constraints, the classical definition must be applied not to the original space S, but to the vastly more complex space S^n whereas the maximum entropy deals only with quantities in S. This simplification is the primary reason for using maximum entropy. However, it is in such cases that the results are least reliable. We shall illustrate with the die experiment. In the absence of prior information, we reach the reasonable conclusion that $p_i = 1/6$. If we know, on the other hand, that the expected value of the zero-one r.v. associated with the event "one" equals 0.1998, say, then the conclusion is that $p_1 = 0.1998$, $p_2 = p_3 = \ldots = p_6 = 0.16004$. Unlike the fair-die case, our trust in the correctness of these values is not great, although we have no other reasonable alternative.

These observations are relevant we believe in the application of the method to spectral estimation problem. In our view, the method is popular not because it leads to an all-pole model as a logical imperative, but rather because the model is numerically simple, and unlike earlier methods, it can detect sharp peaks in the unknown spectra.

Random variables. *Consider a discrete-type r.v. x taking the values* x_i with probability p_i. Clearly, the events $\{x = x_i\}$ form a partition A_x. The entropy of the r.v. x is by definition the entropy of this partition:

$$H(x) = H(A_x) = -\sum_i p_i \ln p_i \qquad (A-7)$$

The entropy of a continuous-type r.v. x cannot be so defined because the events $\{x = x_i\}$ do not form a partition (they are not countable). In this case a limiting argument is used: The r.v. x is approximated by a discrete-type v.v. x_δ taking

the values $x_i = i\delta$ with probability $p_i = f(x_i)\delta$ where $f(x)$ is the density of x. As we see from (A-7)

$$H(x_\delta) = -\sum_i \delta\, f(x_i)\, \ell n\, f(x_i) - \ell n\delta \sum_i \delta\, f(x_i)$$

Hence, $H(x_\delta) \to \infty$ as $\delta \to 0$. This is so because of the underlying assumption that the various values of x can be recognized as distinct no matter how close they are. However,

$$H(x_\delta) + \ell n\delta \to -\int_{-\infty}^{\infty} f(x)\, \ell n\, f(x)\, dx \quad \text{as} \quad \delta \to 0$$

And it is this limit that is used as the entropy of x:

$$H(x) = -\int_{-\infty}^{\infty} f(x)\, \ell n\, f(x)\, dx = -E\{\ell n\, f(x)\} \tag{A-8}$$

The addition of the term $\ell n\delta$ is a recognition of the fact that, in real problems, only values of x whose difference exceeds a certain level can be considered as distinct.

The joint entropy of the vector r.v. $x = (x_1, \ldots, x_r)$ with density $f(x_1, \ldots, x_r)$ is defined similarly:

$$H(x_1, \ldots, x_r) = -E\{\ell n\, f(x_1, \ldots, x_r)\} \tag{A-9}$$

As we know [11], if $y:(y_1, \ldots, y_r)$ is a linear transformation of x, that is, if $y = Ax$ where A is an r by r non-singular matrix, then

$$f(y_1, \ldots, y_r) = \frac{1}{|A|} f(x_1, \ldots, x_r)$$

hence,

$$H(y_1, \ldots, y_r) = -E\{\ell n\, f(x_1, \ldots, x_r)\} + \ell n|A| \tag{A-10}$$

Entropy rate. We shall finally define the entropy rate of a discrete stationary process $x[n]$. From the stationarity of $x[n]$ it follows that the joint entropy

of the r.v. $x[n], \dots, x[n-r+1]$ is independent of n. The entropy rate H_x of the process $x[n]$ is by definition the limit

$$H_x = \frac{1}{r} H(x_1, \dots, x_r) \quad \text{as} \quad r \to \infty \tag{A-12}$$

¶ Suppose that $x[n]$ is the input to a stable causal system with delta response $h[n]$ and system function $H(z)$. If $x[n]$ is applies at $n = -\infty$, then the resulting response $y[n]$ is stationary with entropy rate H_y.

Theorem 1. *If $H(z)$ is minimum phase, then*

$$H_y = H_x + \frac{1}{4\omega_o} \int_{-\omega_o}^{\omega_o} \ell n \left| H(e^{j\omega T}) \right|^2 d\omega \quad \omega_o = \frac{\pi}{T} \tag{A-13}$$

Proof. We can assume, introducing if necessary a change in sign and an appropriate shift of the time-origin, that $h[n] > 0$. If $x[n]$ is applied at $n = 0$, then the resulting response

$$\overline{y}[n] = \sum_{k=0}^{n} x[n-k]\, h[k] \tag{A-14}$$

is not stationary. However, it tends to the stationary process $y[n]$ as $n \to \infty$. Clearly, (A-14) is a linear transformation of the r.v. $x_0 = x[0], \dots, x_n = x[n]$ into the r.v. $y_0 = \overline{y}[0], \dots, y_n = \overline{y}[n]$ of the form $y = Ax$ where

$$A = \begin{bmatrix} h[0] & 0 & \dots & 0 \\ h[1] & h[0] & \dots & 0 \\ \dots & \dots & \dots & \dots \\ h[n] & h[n-1] & & h[0] \end{bmatrix} \qquad |A| = h^{n+1}[0]$$

Hence [see (A-8)]

$$H(y_0, \dots, y_n) = H(x_0, \dots, x_n) + (n+1)\, \ell n\, h[0]$$

Dividing by $n+1$ and making $n \to \infty$, we obtain

$$H_y = H_x + \ell n\, h[0]$$

Therefore, to complete the proof of the theorem it suffices to show that the term $\ell n\, h[0]$ in (A-15) equals the integral in (A-13). Since $|H(e^{j\omega T})|^2 = H(e^{j\omega T})\, H(e^{-j\omega T})$, we conclude with $z = e^{j\omega T}$ that

$$jT\int_{-\omega_o}^{\omega_o} \ell n\,|H(e^{j\omega T})|^2 d\omega = \oint \frac{1}{z}\, \ell n\,[H(z)\, H(1/z)]\, dz$$

where the line integral is along the unit circle. But

$$\oint \frac{1}{z}\, \ell n\, H(z)\, dz = \oint \frac{1}{z}\, \ell n\, H(1/z)\, dz$$

hence it suffices to show that

$$\ell n\, h[0] = \frac{1}{2\pi j} \oint \frac{1}{z}\, \ell n\, H(z)\, dz \qquad \text{(A-16)}$$

From the assumption that $H(z)$ is minimum-phase, it follows that the integrand in (A-16) is analytic for $|z| \geq 1$, hence, the circle of integration can be made arbitrarily large. And since $h[0] = \ell n\, H(z)$ as $z \to \infty$, we conclude that the integral equals

$$\ell n\, h[0] \oint \frac{dz}{z} = 2\pi j\, \ell n\, h[0]$$

and (A-16) results.

Normal processes. If x is a normal r.v. with

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}$$

then $H(x) = -E\{\ell n f(x)\} = \ell n\, \sigma\sqrt{2\pi e}$.

If $\nu[n]$ is normal white noise with $E\{\nu^2[n]\} = \sigma^2$, then

$$f(\nu_1, \cdots, \nu_r) = f(\nu_1) \cdots f(\nu_r)$$

hence

$$H(\nu_1, \cdots, \nu_r) = -E\{\ell n\,[f(\nu_1) \cdots f(\nu_r)]\} = r\, \ell n\, \sigma\sqrt{2\pi e} \qquad \text{(A.17)}$$

From this and (A-13), it follows that if $\nu[n]$ is white noise, then

$$H_\nu = \ln \sigma \sqrt{2\pi e} \tag{A-18}$$

Theorem 2. If $x[n]$ is a normal process with power spectrum $\bar{S}(\omega)$ such that

$$\int_{-\omega_o}^{\omega_o} \ln \bar{S}(\omega)\, d\omega < \infty \tag{A-19}$$

then its entropy rate H_x is given by

$$H_x = \ln \sqrt{2\pi e} + \frac{1}{4\omega_o} \int_{-\omega_o}^{\omega_o} \ln \bar{S}(\omega)\, d\omega \tag{A-20}$$

Proof. Since $x[n]$ is normal, all its statistical properties, including its entropy rate [see (A-12)], can be expressed in terms of its autocorrelation $R[m]$. From this it follows that if another process $y[n]$ has the same autocorrelation $R[m]$, then its entropy rate H_y will equal H_x. Since $\bar{S}(\omega)$ is an even, positive function, and it satisfies the discrete form (A-19) of the Paley-Wiener condition [22], it can be factored into a product [12]

$$\bar{S}(\omega) = H(e^{j\omega T})\, H(e^{-j\omega T}) \tag{A-21}$$

where $H(z)$ is the system function of a real causal minimum phase system. Using as input to this system a white-noise normal process with zero mean and variance one, we obtain as output a normal process $s[n]$ with entropy rate [see (A-13) and (A-18)]

$$H_s = \ln \sqrt{2\pi e} + \frac{1}{4\omega_o} \int_{-\omega_o}^{\omega_o} \ln \left| H(e^{j\omega T}) \right|^2 d\omega$$

and (A-20) follows from (A-21).

Maximum entropy with constraints. The solution of problems involving maximum entropy with constraints in the form of expected values is a simple consequence of the following inequality: If $f(x)$ and $g(x)$ are two arbitrary density

functions, then

$$-\int_{-\infty}^{\infty} f(x) \ln g(x)\, dx \geq -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx \qquad \text{(A-22)}$$

with equality iff $f(x) = g(x)$. Indeed, as it is easy to see, $\ln y \leq 1 - y$ and (A-22) follows readily with $y = g(x)/f(x)$. The above holds also if $f(x)$ and $g(x)$ are replaced by joint densities of any order.

Using (A-22), we shall determine the density $f(x)$ of a r.v. x so as to maximize its entropy $H(x)$ subject to the constraint

$$E\{x^2\} = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2$$

With

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}$$

it follows from (A-22) that

$$-\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx \leq -\int_{-\infty}^{\infty} f(x)\left(-\frac{x^2}{2\sigma^2} - \ln \sigma\sqrt{2\pi}\right) dx = \frac{1}{2} + \ln \sigma\sqrt{2\pi}$$

The left-side equals $H(x)$ and the right-side is specified, hence, $H(x)$ is maximum iff $f(x) = g(x)$, that is, if x is normal with zero mean.

We now wish to find the joint density $f(x_1, \ldots, x_r)$ of the r.v. $x_1, \ldots, x_r$ so as to maximize their entropy $H(x_1, \ldots, x_r)$ subject to the constraints

$$E\{x_i x_j\} = \mu_{ij} \quad i, j = 1, \ldots, r \qquad \text{(A-23)}$$

With

$$x = (x_1, \ldots, x_r) \quad x_t = \begin{bmatrix} x_1 \\ \cdot\cdot \\ x_r \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_{11}, & \cdots, & \mu_{1r} \\ \cdots & \cdots & \cdots \\ \mu_{r1}, & \cdots, & \mu_{rr} \end{bmatrix}$$

and

$$g(x) = \frac{1}{\sqrt{(2\pi)^n |\mu|}} e^{-\frac{1}{2} x \mu^{-1} x_t}$$

we conclude applying the multidimensional form of (A-22) that $f(x) = g(x)$ and

$$H(x) = \ell n \sqrt{(2\pi e)^n |\mu|} \qquad \text{(A-24)}$$

We similarly conclude that if μ_{ij} is specified for $i = j$ only, then $H(x)$ is maximum if the r.v. x_i are normal independent with zero-mean and variance μ_{ii}.

From the above and (A-5) it follows that if $x[n]$ is a random process with given average power $E\{x^2[n]\} = \sigma^2$, then its entropy rate H_x is maximum if $x[n]$ is normal white noise.

18. A. Van Den Bos, "Alternative Interpretation of Maximum Entropy Analysis," IEEE Trans. on Inform. Theory, July 1971.

19. R. N. McDonough, "Maximum Entropy, Spatial Processing of Array Data," Geophys. Vol. 39, December 1974.

20. J. E. Shore and R. W. Johnson, "Axiamatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Trans. Infor. Theory, Vol. IT-26, pp. 26-37, January 1980.

21. J. E. Shore, "Minimum Cross-Entropy Spectral Analysis," IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-29, pp. 230-237, April 1981.

22. A. Papoulis, The Fourier Integral and its Applications, McGraw-Hill, New York, 1962.

REFERENCES

1. E. T. Jaynes, Prior Probabilities, IEEE Trans. Systems Sci. Cybern., Vol. SSC-4, 1968.

2. J. P. Burg, "Maximum Entropy Spectral Analysis," *Intl. Meeting Soc.* of Explor. Geophys., Orla., 1967.

3. R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics, 36, 661-675, 1971.

4. T. J. Ulrych and T. N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geophys. and Space Phys., Vol. 13, February 1975.

5. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, 1979.

6. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, 561-580, April 1975.

7. A. H. Nuttall and G. C. Carter, "A Generalized Framework for Power Spectral Estimation," IEEE Trans. Acoust. Speech Signal Processing, Vol. ASSP-28, pp. 334-335, June 1980.

8. N. Levinson, "The Wiener RMS Error Criterion in Filter Design and Prediction," J. Math. Phys., Vol. 25, No. 4, 1947.

9. J. Makhoul, "Stable and Efficient Lattice Methods for Linear Prediction," IEEE Trans. on ASSP, October 1978.

10. A. H. Nuttall, Naval Underwater Systems Center, Technical Report 5303, March 19:6.

11. A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965

12. A. Papoulis, Signal Analysis, McGraw-Hill, New York, 1977.

13. F. Itakura and S. Saito, "Digital Filtering Techniques for Speech Analysis and Synthesis," presented at the 7th Intl. Congress Acoustics, Budapest, 1971.

14. U. Grenander and G. Szego, Toeplitz Forms and Their Applications, Univ. California Press, 1958.

15. L. D. Davisson, "A Theory of Adaptive Filtering," IEEE Trans. Inform. Theory, Vol. IT-12, pp. 97-102, April 1966.

16. H. Akaike, "A New Look at the Statistical Model Identification," IEEE Trans. Automatic Contr., Vol. AC-19, pp. 716-723, December 1974.

17. D. E. Smylie, G. K. C. Clarice, and T. J. Ulrych, "Analysis of Irregularities in the Earth's Rotation," in Methods of Computational Physics, Vol. 13, pp. 391-430, Academic Press, 1973.

Figure Captions

Fig. 1. Forward predictor filter $\hat{H}_N(z)$.
$\hat{s}_N[n]$: predictor of $s[n]$, $\hat{e}_N[n]$: predictor error.

Fig. 2. Lattice filter.
$\hat{e}_N[n]$: forward error, $\check{e}_N[n]$: backward error.

Fig. 3. Cascade of AR filter $T(z) = 1/\hat{H}_N(z)$ and predictor filter $\hat{H}_N(z)$.

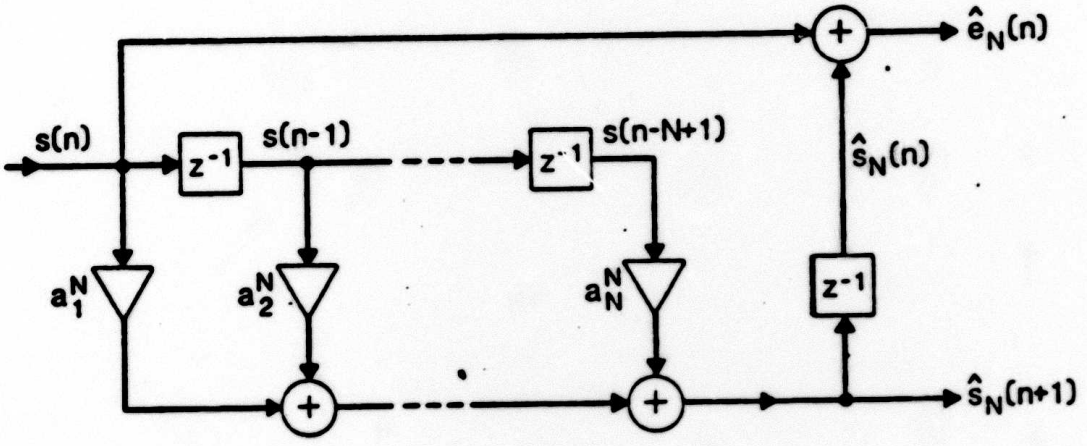


Fig. 1

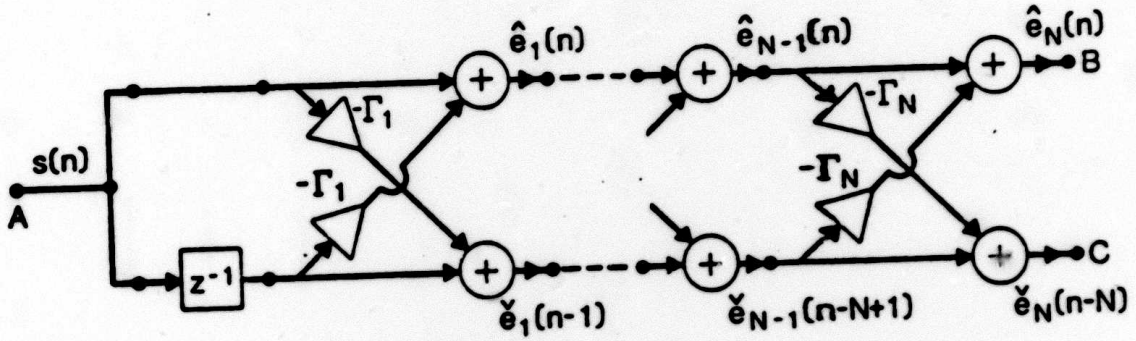


Fig. 2

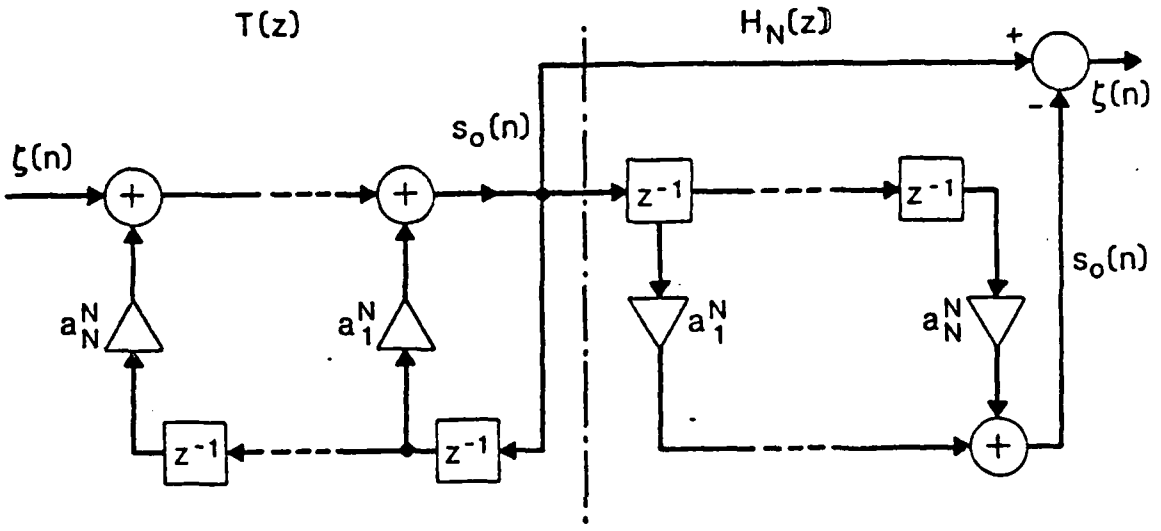


Fig. 3